IE UNIVERSIDAD

TESIS DOCTORAL/
DOCTORAL DISSERTATION

Aprendizaje Automático Responsable e Interpretable para la Gobernanza Sociotécnica en Sistemas de Información: de la Madurez Digital Juvenil a la Democratización de los Datos

Responsible and Interpretable Machine Learning for Socio-Technical Governance in Information Systems: From Youth Digital Maturity to Data Democratization

Aqib Siddiqui

SEGOVIA, 2026

IE UNIVERSIDAD

TESIS DOCTORAL/
DOCTORAL DISSERTATION

Aprendizaje Automático Responsable e Interpretable para la Gobernanza Sociotécnica en Sistemas de Información: de la Madurez Digital Juvenil a la Democratización de los Datos

Responsible and Interpretable Machine Learning for Socio-Technical Governance in Information Systems: From Youth Digital Maturity to Data Democratization

Aqib Siddiqui

Doctoral Thesis Advisor: Konstantina Valogianni

Contents

List of Tables

List of Figures

Acknowledgments

A dissertation can look like a solitary accomplishment from the outside. From the inside, it feels much closer to a long relay race where you keep moving because different people step in at exactly the moments you need them most. This dissertation carries my name, but it was made possible by the time, care, and belief of many others.

My first and deepest thanks go to my supervisor, *Konstantina Valogianni*. You are the reason I began this journey with confidence and the reason I managed to finish it with discipline. You inspired me to take research seriously, to keep aiming for work that is not only rigorous but also high quality and high impact, and to hold myself to standards that I did not always think I could achieve. Your availability, your work ethic, and your unwavering commitment to doing things systematically set the tone for my entire PhD. In the moments when the project felt messy or overwhelming, you helped me return to what mattered and take the next step. I truly could not have completed this dissertation without your guidance and support.

I would like to thank my committee members, *Nancy Pouloudi*, *Alvaro Arenas*, and *Ariana Polyviou*. I am genuinely fortunate to have had scholars of your calibre engage with this dissertation. Beyond the formal role of the committee, I have learned from each of you through collaborations, departmental discussions, seminars and working on research ideas. Your support, feedback and encouragement meant a lot to me and helped me stay motivated.

I am also very grateful to *Laura Maguire*, *Julio de Castro*, and *Maria Muriel*. Thank you for consistently making the programme work in the ways that often remain invisible, answering questions quickly, helping me navigate formal steps, and resolving issues with a calm efficiency that reduced stress at the times it mattered most.

My time at IE also gave me something that mattered more than I expected at the beginning, a circle of academic friends who made the hard stretches easier and the good moments more meaningful. *Vaishnavi*, *Christelle*, and *Prana*, thank you for being there through the full range of this experience: the highs, the lows, and the anxious periods. Your friendship made the PhD feel less like an individual test and more like a shared journey.

This brings me to my family, who supported me with a steadiness that never demanded explanation. To my parents, thank you for your unconditional support, your patience, and the quiet confidence you always showed in me, even when I was uncertain in myself. You made it possible for me to pursue this path with the reassurance that I was never doing it alone. To my sisters, thank you for the constant encouragement, the care, and positivity you brought throughout the process.

Finally, to my wife, *Saania*, thank you for standing by me throughout this process. You have been a constant source of strength, and your patience with the long hours, the distracted mind, and the uneven rhythm of PhD life meant more than I can express. You gave me stability when the work felt unstable, and you made the end of the dissertation feel possible when I was exhausted. Thank you for your love, your support, and for supporting me throughout this journey.

To everyone who helped me reach this point, Thank You!

Abstract

As machine learning becomes embedded in organizational and societal systems, information systems research faces a dual challenge: leveraging ML's analytical power while ensuring that resulting models remain interpretable, feasible, and aligned with human values. This dissertation addresses that challenge through three studies that develop and apply interpretable, cost-aware, and value-sensitive ML approaches to two domains exemplifying contemporary socio-technical complexity: adolescents' digital maturity and European data governance.

The first study proposes an operational assessment system for children's digital maturity that is accurate, interpretable, and feasible for routine use in schools and related settings. Using a cluster-then-predict workflow on multi-country survey data, it discovers distinct maturity profiles and trains supervised models to classify individuals into profiles. To reduce measurement burden without sacrificing decision quality, it introduces a reinforcement learning agent that learns to replicate full-information classifications using only a small subset of the most informative inputs, enabling brief, scalable screening. Model interpretability is integrated through feature attribution methods that identify the competencies most strongly driving predictions, supporting targeted interventions and resource allocation.

The second study examines digital maturity as a developmental process rather than a stable individual trait. It combines explainable machine learning with confirmatory statistical testing to identify which dimensions of digital maturity are most strongly associated with developmental stage, how these relationships change across adolescence, and whether patterns differ by gender. The analysis reveals both foundational capabilities and phase-specific constraints, and documents meaningful within-age heterogeneity, highlighting limits of one-size-fits-all age-based approaches to youth digital governance.

The third study shifts from individual assessment to institutional design, analyzing how values and stakeholder priorities shape large-scale data governance. Using transformer-based natural language processing within a value-sensitive design framework, it maps value emphases across stakeholder consultation letters and the EU Data Act, identifies convergences and tensions, and derives meta-requirements that operationalize data democratization through implementable policy mechanisms.

Methodologically, the dissertation demonstrates how ML-enabled pattern discovery can be combined with confirmatory statistical testing to produce theory-relevant claims without sacrificing inferential rigor. Theoretically, it advances IS research by reframing ML as socio-technical decision infrastructure and by showing how interpretability, information costs, and value alignment can be treated as central design requirements rather than post-hoc considerations. It contributes to IS literature on digital capabilities by conceptualizing youth digital maturity as both profile-based and developmentally dynamic, and it extends IS perspectives on data governance by linking stakeholder value tensions to concrete regulatory mechanisms that enable or constrain data democratization. Practically, the dissertation offers deployable implications: a scalable, low-burden approach for digital maturity screening and intervention targeting in educational and youth-facing contexts, and an evidence-based set of policy meta-requirements that can guide regulators, firms, and civil society in implementing fair access to data while balancing competing values. Taken together, it contributes a coherent view of

ML in information systems as models that are not only predictive tools, but socio-technical instruments that must remain interpretable, operationally feasible, and aligned with human and institutional priorities.

Resumen

A medida que el aprendizaje automático se integra en sistemas organizativos y sociales, la investigación en sistemas de información se enfrenta a un doble desafío: aprovechar la potencia analítica del aprendizaje automático y, al mismo tiempo, garantizar que los modelos resultantes sean interpretables, viables en la práctica y coherentes con valores humanos. Esta tesis doctoral aborda ese desafío mediante tres estudios que desarrollan y aplican enfoques de aprendizaje automático interpretables, sensibles al coste y sensibles a los valores en dos dominios que ejemplifican la complejidad sociotécnica contemporánea: la madurez digital de los adolescentes y la gobernanza europea de los datos.

El primer estudio propone un sistema operativo de evaluación de la madurez digital infantil que es preciso, interpretable y factible para su uso rutinario en centros educativos y contextos afines. A partir de un pipeline de agrupamiento y predicción (cluster-then-predict) sobre datos de encuestas multinacionales, identifica perfiles diferenciados de madurez y entrena modelos supervisados para clasificar a los individuos en dichos perfiles. Para reducir la carga de medición sin sacrificar la calidad de la decisión, introduce un agente de aprendizaje por refuerzo que aprende a reproducir clasificaciones con información completa utilizando solo un subconjunto reducido de las entradas más informativas, lo que permite un cribado breve y escalable. La interpretabilidad se integra mediante métodos de atribución de características que identifican las competencias que impulsan con mayor fuerza las predicciones, apoyando intervenciones focalizadas y una asignación de recursos más eficiente.

El segundo estudio examina la madurez digital como un proceso de desarrollo, en lugar de un rasgo individual estable. Combina aprendizaje automático explicable con contrastes estadísticos confirmatorios para identificar qué dimensiones de la madurez digital se asocian con mayor intensidad a la etapa evolutiva, cómo cambian estas relaciones a lo largo de la adolescencia y si los patrones difieren por género. El análisis revela tanto capacidades fundacionales como restricciones específicas por fase, y documenta una heterogeneidad sustantiva dentro de cada grupo de edad, subrayando los límites de los enfoques uniformes basados exclusivamente en la edad para la gobernanza digital juvenil.

El tercer estudio desplaza el foco desde la evaluación individual hacia el diseño institucional, analizando cómo los valores y las prioridades de los grupos de interés configuran la gobernanza de los datos a gran escala. Mediante procesamiento del lenguaje natural basado en transformadores, integrado en un marco de diseño sensible a los valores, cartografía los énfasis valorativos en cartas de consulta de actores relevantes y en la Ley de Datos de la Unión Europea, identifica convergencias y tensiones, y deriva meta-requisitos que operacionalizan la democratización de los datos mediante mecanismos de política implementables.

Desde el punto de vista metodológico, la tesis demuestra cómo el descubrimiento de patrones habilitado por aprendizaje automático puede combinarse con pruebas estadísticas confirmatorias para producir afirmaciones relevantes para la teoría sin perder rigor inferencial. En términos teóricos, contribuye a la investigación en sistemas de información al replantear el aprendizaje automático como infraestructura sociotécnica de decisión y al mostrar que la interpretabilidad, los costes de información y la alineación con valores pueden tratarse como requisitos centrales de diseño, y no como consideraciones posteriores. Aporta a la

literatura sobre capacidades digitales al conceptualizar la madurez digital juvenil como simultáneamente basada en perfiles y dinámicamente evolutiva, y amplía las perspectivas de sistemas de información sobre gobernanza de datos al vincular tensiones valorativas de los actores con mecanismos regulatorios concretos que habilitan o restringen la democratización de los datos. En el plano práctico, la tesis ofrece implicaciones desplegables: un enfoque escalable y de baja carga para el cribado de madurez digital y la focalización de intervenciones en contextos educativos y orientados a jóvenes, y un conjunto de meta-requisitos, sustentados empíricamente, que pueden guiar a reguladores, empresas y sociedad civil en la implementación de un acceso justo a los datos equilibrando valores en conflicto. En conjunto, la tesis propone una visión coherente del aprendizaje automático en sistemas de información como modelos que no son solo herramientas predictivas, sino instrumentos sociotécnicos que deben seguir siendo interpretables, viables en términos operativos y alineados con prioridades humanas e institucionales.

Introduction

The information systems (IS) discipline is fundamentally concerned with understanding how digital technologies are designed, adopted, and embedded in organizational and societal contexts (Orlikowski and Iacono, 2001; Benbasat and Zmud, 2003). Central to this view is recognizing that information technologies are not neutral instruments but socio-technical artifacts whose effects depend on how they are configured, interpreted, and governed in practice. As the nature of these technologies has evolved from transaction processing systems to enterprise platforms and to algorithmic decision systems, so has the challenge of specifying what responsible design entails. Machine learning, now widely deployed in classification, prediction, and decision support, represents a significant moment in this evolution as these systems learn from data, operate probabilistically, and participate in consequential organizational processes in ways that raise distinct questions about interpretability, feasibility, and value alignment (Padmanabhan et al., 2022; Berente et al., 2021).

The Algorithmic Turn and the Evolving IT Artifact

IS research has repeatedly been shaped by shifts in the nature of the IT artifact and its organizational and societal consequences (Orlikowski and Iacono, 2001; Benbasat and Zmud, 2003; Akhlaghpour et al., 2013). Foundational work in IS emphasizes that IS scholarship is not simply about technology, but about the socio-technical phenomena that arise when digital artifacts are designed, used, institutionalized, and contested in practice (Sidorova et al., 2008). Over the past decades, two developments have intensified this agenda. First, digital infrastructures like platforms, mobile devices, connected products, and cloud services have generated high-volume and increasingly granular data about behavior, interaction, and organizational process (Tilson et al., 2010). Second, algorithmic systems have become embedded in core organizational activities including classification, recommendation, resource allocation, and compliance (Faraj et al., 2018). As a result, many contemporary IT artifacts differ from traditional software systems. Rather than being fully specified by deterministic rules, they increasingly rely on learning-based systems that produce probabilistic outputs and update through exposure to data.

Machine learning (ML) sits at the center of this shift. ML refers to statistical and computational methods that improve task performance through data and experience (Mitchell, 1997; Jordan and Mitchell, 2015). In organizational settings, ML systems do not merely summarize information but they participate in decision processes by generating categories, scores, rankings, predictions, or recommendations that shape attention and action. This matters for IS research because the legitimacy and effectiveness of such systems often depend not only on predictive performance, but also on whether outputs can be understood, contested, and operationalized within institutional constraints and governance requirements (Doshi-Velez and Kim, 2017; Selbst et al., 2019). The algorithmic turn thus creates both empirical opportunity and an evaluation challenge on how should ML-based artifacts be assessed when they function as part of organizational decision-making and governance?

This dissertation addresses that challenge by examining ML not only as a statistical method, but as a socio-technical decision support tool embedded in specific institutional contexts. It argues that for ML to contribute meaningfully to IS research and practice, it must be designed to be interpretable, feasible under real information and measurement constraints, and aligned with institutional values. Across three studies in two domains(adolescents' Digital Maturity and European Data Governance) the dissertation develops ML-based artifacts and solutions that explicitly incorporate these requirements into their design and evaluation.

Machine Learning in IS: From Method to Object of Study to Design Artifact

The use of ML in IS has grown substantially, prompting renewed editorial and theoretical attention to what constitutes a strong ML contribution in the discipline (Padmanabhan et al., 2022; Berente et al., 2021). A useful way to position this dissertation is to distinguish three complementary orientations in the literature.

ML as a Methodological Toolkit

A first orientation treats ML primarily as a methodological toolkit for prediction and pattern discovery, often motivated by the distinction between explaining and predicting (Shmueli, 2010; Shmueli and Koppius, 2011; Breiman, 2001). Within this view, ML is valuable because it can represent complex, nonlinear relationships and perform well in high-dimensional settings. Predictive work can also be theory-relevant when it identifies robust patterns that invite explanation or inform decision support (Shmueli and Koppius, 2011). This orientation has been productive for IS, but it can under-specify the socio-technical conditions under which predictive artifacts become usable and legitimate in practice (Selbst et al., 2019). A model optimized for predictive accuracy on held-out data may nonetheless fail when deployed if stakeholders cannot interpret its outputs, if required input data are unavailable or costly to collect, or if the system's recommendations conflict with institutional norms and governance requirements.

ML as An Object of Social and Organizational Inquiry

A second orientation treats ML and AI systems as objects of study, focusing on adoption, trust, organizational consequences, and ethical implications. Research on algorithmic advice has documented both algorithm aversion and algorithm appreciation in human judgment contexts (Dietvorst et al., 2015; Logg et al., 2019). IS literature has also explored how learning algorithms reshape work and organizing and raise questions of accountability, fairness, and governance (Faraj et al., 2018; Martin, 2019; Robert et al., 2020). This orientation emphasizes that ML is socially embedded, but it often treats the algorithmic system as a fixed black box rather than as a malleable artifact whose internal design can be adapted to socio-technical requirements. Studies of algorithmic bias, for instance, have illuminated how ML systems can encode and perpetuate discrimination (Barocas and Selbst, 2016), yet such work typically analyzes deployed systems rather than specifying how alternative design choices might mitigate these harms. The result is a critical literature strong in diagnosis but limited in prescriptive guidance for those who must build and configure these systems.

ML as a Socio-Technical Design Artifact

The primary aim of this dissertation is to foreground a third orientation, that is, ML as a design artifact that can and should be engineered to meet socio-technical requirements. IS literature has long recognized that the value of an artifact depends not only on its technical capabilities but also on its fit with the organizational, informational, and institutional environment in which it operates (Sarker et al., 2019; Hevner et al., 2004). Extending that logic to ML implies that model design choices like what data is required, how decisions are produced, how outputs are explained, what constraints are respected, and which values are prioritized, etc are not secondary implementation details. They are constitutive of the artifact itself. Under this perspective, performance becomes multidimensional. An ML system can be highly accurate and still fail as an IS artifact if it is too costly to operate, too opaque to support responsible action, or misaligned with institutional values and governance obligations (Padmanabhan et al., 2022; Selbst et al., 2019).

Design Requirements for Responsible ML Artifacts in Information Systems

Building on this design-artifact outlook, the dissertation focuses on three requirements that recurrently determine whether ML systems can function as decision support tools in socio-technical settings. They are interpretability (Rudin, 2019), feasibility under information constraints (Padmanabhan et al., 2022), and value sensitivity (Friedman et al., 2006).

Interpretability and Explainability: Making ML Usable and Contestable

Many high-performing ML models are complex, and complexity can reduce transparency for stakeholders who must understand, contest, or operationalize model outputs (Lipton, 2018). The interpretability literature emphasizes that interpretability should be specified relative to audience and purpose, rather than treated as a universal property (Doshi-Velez and Kim, 2017). It is therefore useful to distinguish interpretability from explainability.

Interpretability refers to models whose decision logic can be understood directly from their structure, for example, through sparse linear models or shallow decision trees. Explainability refers to post-hoc techniques that approximate or summarize model behavior for specific predictions, particularly for complex models. This dissertation uses SHAP (SHapley Additive exPlanations) as a primary explanation method because it provides theoretically grounded, consistent feature-attribution summaries that can support stakeholder-facing reasoning (Lundberg and Lee, 2017). At the same time, the dissertation follows the caution that post-hoc explanations do not automatically resolve accountability concerns in consequential settings. They should be treated as part of a broader design strategy rather than as a substitute for responsible system design (Rudin, 2019).

Cost-Awareness and Operational Feasibility Under Measurement Constraints

In many IS contexts, measurement is not free. Data collection consumes time, attention, and institutional capacity and may create privacy, compliance, or legitimacy burdens. This motivates modeling strategies that treat information acquisition as a scarce resource and optimize the trade-off between decision quality and measurement burden (Janisch et al., 2020). Reinforcement learning (RL) is particularly useful here because it provides a formal language for sequential decision-making under constraints and can learn policies that balance accuracy against cost through interaction (Sutton and Barto, 2018). When ML outputs are intended to support routine decisions in resource-constrained settings, feasibility becomes an integral component of model performance.

Value Sensitivity and Governance

ML systems embed normative choices about what counts as performance, what risks are tolerated, and which stakeholders benefit. Scholars have emphasized that these choices are not neutral and that fairness, accountability, and transparency concerns should be addressed as part of system design and governance (Selbst et al., 2019; Mitchell et al., 2021). Value Sensitive Design (VSD) provides a conceptual approach for treating values as part of system specification rather than as an afterthought (Friedman et al., 2006, 2019). In policy contexts, these issues intersect with legal and institutional arrangements, including European data governance initiatives that structure data access, portability, safeguards, and accountability (European Parliament and Council of the European Union, 2016; European Union, 2023). A key claim of this dissertation is that values become actionable in socio-technical systems when they are linked to mechanisms—concrete design and governance features that allocate rights, obligations, safeguards, and recourse.

Research Question and Dissertation Contribution

The dissertation is organized around one overarching question:

> *How can machine learning systems be designed as socio-technical decision support artifacts that remain interpretable, operationally feasible under real information constraints, and aligned with institutional values?*

The dissertation contributes to IS research by advancing a design-oriented evaluation of ML artifacts. Specifically, it, *(i)* expands evaluation beyond predictive accuracy to include interpretability, feasibility under measurement constraints, and value sensitivity as core design requirements; *(ii)* demonstrates a concrete design pattern for cost-aware decision support in settings where information collection is expensive or burdensome; *(iii)* articulates and implements a disciplined discovery-confirmation workflow in which ML supports pattern discovery while confirmatory modeling supports defensible inference; and *(iv)* extends value-sensitive analysis to techno-regulatory design by linking stakeholder value tensions to implementable governance mechanisms and meta-requirements.

The dissertation comprises three studies, each developing an ML artifact tailored to a distinct socio-technical requirement.

Chapter 1: Cost-Aware Screening for Adolescents' Digital Maturity

Chapter 1 develops an operational approach to assessing adolescents' digital maturity designed for youth-facing settings where measurement burden is a binding constraint. The chapter uses a cluster-then-predict workflow to discover maturity profiles and train an ensemble classifier (combining Random Forest and Deep Neural Network) that classifies individuals into those profiles. It then acknowledges feature acquisition as a budgeted decision problem and trains a Deep Reinforcement Learning (DRL) policy to approximate full-information classifications using only a small subset of the most informative survey items (Janisch et al., 2020; Sutton and Barto, 2018). SHAP explanations are used to identify which competencies drive predictions and to connect classifications to intervention-relevant reasoning (Lundberg and Lee, 2017). The resulting artifact is evaluated not only by accuracy, but also by its feasibility for routine screening in educational and social work contexts.

Chapter 2: Interpretable discovery with confirmatory discipline

Chapter 2 treats digital maturity as developmentally dynamic rather than as a stable trait. It addresses a common challenge in ML-enabled social science, flexible models can reveal complex patterns, but without careful design such patterns can be difficult to treat as inferential knowledge. The chapter therefore adopts a two-stage discovery-confirmation workflow using data from 3,018 adolescents. In the discovery stage, Random Forest models and SHAP are used to identify the maturity dimensions most associated with developmental stage and how these relationships vary across adolescence and by gender (Lundberg and Lee, 2017). In the confirmation stage, the candidate relationships are tested using semiparametric regression models designed for transparent statistical evaluation, thereby guarding against over-interpretation (Shmueli, 2010). This chapter demonstrates how interpretability can support theory-relevant insight while preserving inferential accountability.

Chapter 3: Value-sensitive ML for European data governance

Chapter 3 shifts the unit of analysis from individual assessment to institutional and policy design. It examines how values and stakeholder priorities shape data governance and data democratization in the European Union, focusing on 102 stakeholder consultation letters submitted during the development of the EU Data Act. Treating regulation as a socio-technical design artifact, the chapter combines transformer-based NLP (EUBERT, a domain-adapted model for EU legal and policy text) with a VSD-informed analytic frame to map value emphases, identify tensions across stakeholder groups, and derive meta-requirements that connect values to implementable mechanisms, including access rights, portability provisions, safeguards, interoperability obligations, and accountability structures (Friedman et al., 2006, 2019; European Union, 2023). The result is an approach that uses ML to support transparency about normative trade-offs while remaining traceable to concrete governance design choices.

Together, the dissertation advances a coherent view of responsible ML in IS. ML systems should be evaluated as socio-technical decision artifacts whose legitimacy and usefulness depend on interpretability, feasibility, and value alignment, not only on predictive performance. The following chapters develop this argument empirically across adolescent digital maturity and European data governance, while illustrating design patterns that are transferable to other high-stakes IS settings where algorithmic systems mediate human decisions.

Introducción

La disciplina de Sistemas de Información (SI) se ocupa, de manera fundamental, de comprender cómo se diseñan, adoptan e integran las tecnologías digitales en contextos organizativos y sociales (Orlikowski and Iacono, 2001; Benbasat and Zmud, 2003). Un elemento central de esta perspectiva es reconocer que las tecnologías de la información no son instrumentos neutrales, sino artefactos socio-técnicos cuyos efectos dependen de cómo se configuran, se interpretan y se gobiernan en la práctica. A medida que la naturaleza de estas tecnologías ha evolucionado desde sistemas de procesamiento de transacciones hasta plataformas empresariales y, más recientemente, sistemas algorítmicos de decisión, también se ha intensificado el reto de especificar qué implica un diseño responsable. El aprendizaje automático (Machine Learning, ML), hoy ampliamente desplegado en tareas de clasificación, predicción y apoyo a la decisión, constituye un momento clave en esta evolución: estos sistemas aprenden a partir de datos, operan de forma probabilística y participan en procesos organizativos con consecuencias relevantes, lo que plantea preguntas específicas sobre interpretabilidad, viabilidad y alineación con valores (Padmanabhan et al., 2022; Berente et al., 2021).

El giro algorítmico y la evolución del artefacto TI

La investigación en SI ha estado reiteradamente marcada por cambios en la naturaleza del artefacto TI y por sus consecuencias organizativas y sociales (Orlikowski and Iacono, 2001; Benbasat and Zmud, 2003; Akhlaghpour et al., 2013). Trabajos fundacionales subrayan que el objeto de SI no es la tecnología en abstracto, sino los fenómenos socio-técnicos que emergen cuando los artefactos digitales se diseñan, se usan, se institucionalizan y se disputan en la práctica (Sidorova et al., 2008). En las últimas décadas, dos desarrollos han intensificado esta agenda. En primer lugar, infraestructuras digitales como plataformas, dispositivos móviles, productos conectados y servicios en la nube han generado datos de gran volumen y trazas cada vez más granulares sobre comportamiento, interacción y procesos organizativos (Tilson et al., 2010). En segundo lugar, los sistemas algorítmicos se han incrustado en actividades nucleares de las organizaciones, incluyendo clasificación, recomendación, asignación de recursos y cumplimiento normativo (Faraj et al., 2018). Como resultado, muchos artefactos TI contemporáneos difieren de los sistemas de software tradicionales: en lugar de estar plenamente especificados por reglas deterministas, dependen crecientemente de sistemas basados en aprendizaje que producen salidas probabilísticas y se actualizan mediante exposición a datos.

El ML se sitúa en el centro de este cambio. El ML se refiere a métodos estadísticos y computacionales que mejoran el rendimiento en una tarea a través de datos y experiencia (Mitchell, 1997; Jordan and Mitchell, 2015). En entornos organizativos, los sistemas de ML no se limitan a resumir información; participan en procesos de decisión generando categorías, puntuaciones, rankings, predicciones o recomendaciones que orientan la atención y la acción. Esto es relevante para SI porque la legitimidad y la efectividad de estos sistemas suelen depender no solo del rendimiento predictivo, sino también de si sus salidas pueden comprenderse, cuestionarse y operacionalizarse dentro de restricciones institucionales y requisitos de gobernanza (Doshi-Velez and Kim, 2017; Selbst et al., 2019). El giro algorítmico, por tanto, crea oportunidades empíricas y, al mismo tiempo, un desafío de evaluación: *¿cómo deben evaluarse los artefactos basados en ML cuando funcionan como parte de la toma de decisiones organizativa y de la gobernanza?*

Esta tesis aborda ese desafío examinando el ML no solo como método estadístico, sino como herramienta de apoyo a la decisión socio-técnica incrustada en contextos institucionales específicos. Sostiene que, para que el ML contribuya de manera significativa a la investigación y la práctica en SI, debe diseñarse de forma interpretable, ser viable bajo restricciones reales de información y medición, y estar alineado con valores institucionales. A través de tres estudios en dos dominios (madurez digital adolescente y gobernanza europea de datos), la tesis desarrolla artefactos y soluciones basados en ML que incorporan explícitamente estos requisitos en su diseño y evaluación.

El aprendizaje automático en SI: de método, a objeto de estudio, a artefacto de diseño

El uso de ML en SI ha crecido de forma sustantiva, lo que ha reactivado la atención editorial y teórica sobre qué constituye una contribución sólida basada en ML dentro de la disciplina (Padmanabhan et al., 2022; Berente et al., 2021). Una manera útil de situar esta tesis es distinguir tres orientaciones complementarias en la literatura.

ML como herramienta metodológica

Una primera orientación trata el ML principalmente como un conjunto de herramientas metodológicas para la predicción y el descubrimiento de patrones, motivada a menudo por la distinción entre explicar y predecir (Shmueli, 2010; Shmueli and Koppius, 2011; Breiman, 2001). Desde esta perspectiva, el ML es valioso porque puede representar relaciones complejas y no lineales y rendir bien en entornos de alta dimensionalidad. El trabajo predictivo puede ser relevante para la teoría cuando identifica patrones robustos que invitan a explicación o informan el apoyo a la decisión (Shmueli and Koppius, 2011). Esta orientación ha sido productiva para SI, pero puede infraespecificar las condiciones socio-técnicas bajo las cuales los artefactos predictivos resultan utilizables y legítimos en la práctica (Selbst et al., 2019). Un modelo optimizado para precisión predictiva en datos de validación puede fracasar al desplegarse si los actores no pueden interpretar sus salidas, si los datos de entrada son inaccesibles o costosos de recopilar, o si las recomendaciones entran en conflicto con normas institucionales y requisitos de gobernanza.

ML como objeto de indagación social y organizativa

Una segunda orientación trata los sistemas de ML y de IA como objetos de estudio, poniendo el foco en adopción, confianza, consecuencias organizativas e implicaciones éticas. La investigación sobre asesoramiento algorítmico ha documentado tanto aversión a los algoritmos como apreciación algorítmica en contextos de juicio humano (Dietvorst et al., 2015; Logg et al., 2019). La literatura de SI también ha analizado cómo los algoritmos de aprendizaje reconfiguran el trabajo y la organización y plantean cuestiones de rendición de cuentas, equidad y gobernanza (Faraj et al., 2018; Martin, 2019; Robert et al., 2020). Esta orientación subraya que el ML está socialmente incrustado, pero con frecuencia trata el sistema algorítmico como una caja negra fija, más que como un artefacto maleable cuyo diseño interno puede adaptarse a requisitos socio-técnicos. Por ejemplo, los estudios sobre sesgo algorítmico han mostrado cómo los sistemas de ML pueden codificar y perpetuar discriminación (Barocas and Selbst, 2016), pero suelen analizar sistemas desplegados en lugar de especificar cómo decisiones de diseño alternativas podrían mitigar dichos daños. El resultado es una literatura crítica fuerte en diagnóstico, pero más limitada en orientación prescriptiva para quienes deben construir y configurar estos sistemas.

ML como artefacto de diseño socio-técnico

El objetivo principal de esta tesis es enfatizar una tercera orientación: el ML como artefacto de diseño que puede y debe ser ingenierizado para cumplir requisitos socio-técnicos. La literatura de SI reconoce desde hace tiempo que el valor de un artefacto depende no solo de sus capacidades técnicas, sino también de su ajuste con el entorno organizativo, informacional e institucional en el que opera (Sarker et al., 2019; Hevner et al., 2004). Extender esta lógica al ML implica que decisiones de diseño como qué datos se requieren, cómo se producen las decisiones, cómo se explican las salidas, qué restricciones se respetan y qué valores se priorizan, no son detalles de implementación secundarios. Son constitutivos del propio artefacto. Desde esta perspectiva, el rendimiento se vuelve multidimensional. Un sistema de ML puede ser muy preciso y, aun así, fallar como artefacto de SI si es demasiado costoso de operar, demasiado opaco para apoyar una acción responsable o está desalineado con valores institucionales y obligaciones de gobernanza (Padmanabhan et al., 2022; Selbst et al., 2019).

Requisitos de diseño para artefactos responsables de ML en SI

Sobre esta perspectiva orientada al artefacto, la tesis se centra en tres requisitos que de forma recurrente determinan si los sistemas de ML pueden funcionar como herramientas de apoyo a la decisión en entornos socio-técnicos: interpretabilidad (Rudin, 2019), viabilidad bajo restricciones de información (Padmanabhan et al., 2022) y sensibilidad a valores (Friedman et al., 2006).

Interpretabilidad y explicabilidad: hacer el ML utilizable y cuestionable

Muchos modelos de ML con alto rendimiento son complejos, y la complejidad puede reducir la transparencia para actores que deben comprender, impugnar u operacionalizar sus salidas (Lipton, 2018). La literatura sobre interpretabilidad enfatiza que esta debe especificarse en relación con la audiencia y el propósito, en lugar de tratarse como una propiedad universal (Doshi-Velez and Kim, 2017). Por ello, resulta útil distinguir interpretabilidad de explicabilidad.

La interpretabilidad se refiere a modelos cuya lógica de decisión puede entenderse directamente a partir de su estructura (por ejemplo, modelos lineales dispersos o árboles de decisión poco profundos). La explicabilidad se refiere a técnicas post hoc que aproximan o resumen el comportamiento del modelo para predicciones específicas, especialmente en modelos complejos. Esta tesis utiliza SHAP (SHapley Additive exPlanations) como método principal de explicación, porque proporciona resúmenes de atribución de características teóricamente fundamentados y consistentes que pueden apoyar razonamientos orientados a actores no técnicos (Lundberg and Lee, 2017). Al mismo tiempo, la tesis asume la cautela de que las explicaciones post hoc no resuelven automáticamente problemas de rendición de cuentas en contextos con consecuencias relevantes: deben entenderse como parte de una estrategia de diseño más amplia, no como sustituto del diseño responsable del sistema (Rudin, 2019).

Sensibilidad al coste y viabilidad operativa bajo restricciones de medición

En muchos contextos de SI, medir no es gratis. La recopilación de datos consume tiempo, atención y capacidad institucional, y puede generar cargas de privacidad, cumplimiento o legitimidad. Esto motiva estrategias de modelización que tratan la adquisición de información como un recurso escaso y optimizan el compromiso entre calidad de decisión y carga de medición (Janisch et al., 2020). El aprendizaje por refuerzo (Reinforcement Learning, RL) es especialmente útil porque ofrece un lenguaje formal para decisiones secuenciales bajo restricciones y permite aprender políticas que equilibran precisión y coste a través de interacción (Sutton and Barto, 2018). Cuando las salidas del ML pretenden apoyar decisiones rutinarias en entornos con recursos limitados, la viabilidad se convierte en un componente integral del rendimiento del modelo.

Sensibilidad a valores y gobernanza

Los sistemas de ML incorporan elecciones normativas sobre qué cuenta como rendimiento, qué riesgos se toleran y qué actores se benefician. Diversos autores han subrayado que estas decisiones no son neutrales y que las preocupaciones de equidad, rendición de cuentas y transparencia deben abordarse como parte del diseño y la gobernanza (Selbst et al., 2019; Mitchell et al., 2021). El enfoque de Value Sensitive Design (VSD) ofrece un marco conceptual para tratar los valores como parte de la especificación del sistema, y no como un elemento posterior (Friedman et al., 2006, 2019). En contextos de política pública, estas cuestiones se cruzan con arreglos legales e institucionales, incluyendo iniciativas europeas de gobernanza de datos que estructuran acceso, portabilidad, salvaguardas y responsabilidad (European Parliament and Council of the European Union, 2016; European Union, 2023). Un argumento central de esta tesis es que los valores se vuelven accionables en sistemas socio-técnicos cuando se vinculan a mecanismos: rasgos concretos de diseño y gobernanza que asignan derechos, obligaciones, salvaguardas y vías de recurso.

Pregunta de investigación y contribución de la tesis

La tesis se articula en torno a una pregunta general:

> *¿Cómo pueden diseñarse sistemas de aprendizaje automático como artefactos socio-técnicos de apoyo a la decisión que permanezcan interpretables, operativamente viables bajo restricciones reales de información y alineados con valores institucionales?*

La tesis contribuye a la investigación en SI al avanzar una evaluación orientada al diseño de artefactos de ML. En concreto, *(i)* amplía la evaluación más allá de la precisión predictiva para incluir interpretabilidad, viabilidad bajo restricciones de medición y sensibilidad a valores como requisitos nucleares de diseño; *(ii)* demuestra un patrón de diseño concreto para apoyo a la decisión sensible al coste en contextos donde recopilar información es caro o gravoso; *(iii)* articula e implementa un flujo de trabajo disciplinado de descubrimiento y confirmación, en el que el ML apoya el descubrimiento de patrones mientras la modelización confirmatoria sustenta una inferencia defendible; y *(iv)* extiende el análisis sensible a valores al diseño tecno-regulatorio al vincular tensiones de valores de los actores con mecanismos de gobernanza implementables y con meta-requisitos.

La tesis consta de tres estudios, cada uno desarrollando un artefacto de ML adaptado a un requisito socio-técnico distinto.

Capítulo 1: cribado sensible al coste para la madurez digital adolescente

El Capítulo 1 desarrolla un enfoque operativo para evaluar la madurez digital adolescente, diseñado para entornos orientados a jóvenes donde la carga de medición es una restricción vinculante. El capítulo utiliza un flujo de trabajo de *cluster-then-predict* para descubrir perfiles de madurez y entrenar un clasificador en ensamblado (combinando Random Forest y una red neuronal profunda) que asigna individuos a dichos perfiles. A continuación, reconoce la adquisición de características como un problema de decisión con presupuesto y entrena una política de Deep Reinforcement Learning (DRL) para aproximar clasificaciones con información completa utilizando solo un subconjunto reducido de ítems altamente informativos (Janisch et al., 2020; Sutton and Barto, 2018). Se emplean explicaciones SHAP para identificar qué competencias impulsan las predicciones y para conectar las clasificaciones con razonamientos relevantes para la intervención (Lundberg and Lee, 2017). El artefacto resultante se evalúa no solo por precisión, sino también por su viabilidad para un cribado rutinario en contextos educativos y de intervención socioeducativa.

Capítulo 2: descubrimiento interpretable con disciplina confirmatoria

El Capítulo 2 trata la madurez digital como un fenómeno dinámico del desarrollo y no como un rasgo estable. Aborda un reto habitual en ciencias sociales habilitadas por ML: los modelos flexibles pueden revelar patrones complejos, pero sin un diseño cuidadoso dichos patrones son difíciles de tratar como conocimiento inferencial. Por ello, el capítulo adopta un flujo de trabajo en dos etapas (descubrimiento y confirmación) con datos de 3.018 adolescentes. En la etapa de descubrimiento, se utilizan modelos Random Forest y SHAP para identificar qué dimensiones de madurez se asocian más con la etapa del desarrollo y cómo varían estas relaciones a lo largo de la adolescencia y por género (Lundberg and Lee, 2017). En la etapa de confirmación, las relaciones candidatas se contrastan mediante modelos de regresión semiparamétrica diseñados para una evaluación estadística transparente, evitando la sobreinterpretación (Shmueli, 2010). Este capítulo muestra cómo la interpretabilidad puede producir conocimiento relevante para la teoría sin renunciar a la rendición de cuentas inferencial.

Capítulo 3: ML sensible a valores para la gobernanza europea de datos

El Capítulo 3 desplaza la unidad de análisis desde la evaluación individual hacia el diseño institucional y de políticas públicas. Examina cómo los valores y las prioridades de los actores influyen en la gobernanza de datos y en la democratización de datos en la Unión Europea, centrándose en 102 cartas de consulta presentadas durante el desarrollo del Reglamento de Datos (EU Data Act). Al tratar la regulación como artefacto socio-técnico de diseño, el capítulo combina PLN basada en transformadores (EUBERT, un modelo adaptado al dominio para texto jurídico y de políticas de la UE) con un marco analítico inspirado en VSD para mapear énfasis de valores, identificar tensiones entre grupos de actores y derivar meta-requisitos que conectan valores con mecanismos implementables, incluyendo derechos de acceso, disposiciones de portabilidad, salvaguardas, obligaciones de interoperabilidad y estructuras de rendición de cuentas (Friedman et al., 2006, 2019; European Union, 2023). El resultado es un enfoque que utiliza ML para apoyar la transparencia sobre compromisos normativos, manteniendo la trazabilidad hacia decisiones concretas de diseño de gobernanza.

En conjunto, la tesis desarrolla una visión coherente del ML responsable en SI. Los sistemas de ML deben evaluarse como artefactos socio-técnicos de decisión cuya legitimidad y utilidad dependen de interpretabilidad, viabilidad y alineación con valores, y no únicamente del rendimiento predictivo. Los capítulos siguientes desarrollan este argumento de forma empírica en los dominios de madurez digital adolescente y gobernanza europea de datos, y proponen patrones de diseño transferibles a otros entornos de SI con consecuencias relevantes donde los sistemas algorítmicos median decisiones humanas.

Interpretable and Cost-Aware Assessment of Children's Digital Maturity via Deep Reinforcement Learning

Abstract

Children's digital maturity, defined by the skills, self-regulation, and digital awareness that enable technology use to support rather than harm their well-being, is challenging to measure in ways that are accurate, interpretable, and feasible for routine use in schools and other similar setups. We address this problem in two parts. First, we discover maturity profiles and build a classifier to assign students to them. Second, we deliver the same decisions with minimal information so screening can be brief and cost effective. Methodologically, we use a cluster-then-predict pipeline. Unsupervised machine learning first identifies maturity profiles of children. We then train a two-model ensemble that combines a tree-based classifier and a neural-network to improve predictive accuracy. Interpretability is built in with SHAP attributions that indicate which competencies most influence each prediction. To make this assessment short and cost effective, we first reduce inputs to the most informative features identified on the training data. We then introduce a deep reinforcement learning policy trained with teacher distillation in which during training the policy learns from the full-information DNN's probabilities and at deployment it uses only the top three DIMI items identified on the training data to assign a profile in a single step. We calibrate and test the approach using survey responses from Austria, Germany, Greece, and Denmark (N = 1,440 adolescents). The combined classifier predicts maturity profiles with high accuracy. The minimal-information policy matches or slightly exceeds the full-information benchmark while reducing inputs by two thirds. SHAP analyses highlight different primary levers by country, for example risk awareness and digital literacy in Austria, respect toward others and support seeking in Germany, autonomy of choice and impulse regulation in Greece, and autonomy within digital contexts in Denmark. The results indicate three practical uses. Schools can run brief, interpretable screens and focus on the few competencies with the largest payoffs. Technology platforms can trigger safety-by-design features when those competencies are weak. Policymakers can target funding to the skills that matter most in each context. In addition, ministries and NGOs can use the same indicators to prioritize outreach and training where low-maturity profiles are most prevalent.[1]

[1]This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 870578.

Key Words: Digital Maturity, Adolescent ICT Use, Interpretable Machine Learning, Deep Reinforcement Learning, Educational Operations

1.1 Introduction

The rapid integration of digital technologies into everyday life has deeply influenced how adolescents grow, develop and interact. Today's youth, often known as the 'digital natives' (Prensky, 2001), spend considerable time in an online environment that is deeply intertwined with their offline life. This reality presents both extraordinary opportunities and significant challenges. Adolescents use digital platforms for self-directed learning, social connection, and creative expression (Jackson et al., 2008; Livingstone, 2012; Frauenberger et al., 2011), with Information and Communication Technologies (ICTs) and internet access rates approaching ubiquity across Europe, USA and Australia (Eurostat, 2019; Rideout and Robb, 2020). However, the same technologies also introduce risks associated with excessive use, disordered attention, and compromised mental well-being, risks that were amplified during periods of intensified online engagement such as the COVID-19 pandemic (Allen et al., 2019; Odgers and Jensen, 2020; Pandya and Lodha, 2021). Ultimately, access to digital technologies is no longer the key issue. Instead, ensuring that such engagement supports, rather than undermines, youth well-being is paramount (Livingstone et al., 2018).

Despite growing awareness of these dualities, the bulk of research and policy focus remains centered on digital literacy, educational outcomes, social media use or on attributes like reading (Kerckaert et al., 2015; Hu and Yu, 2021; Luo et al., 2020) or behavioral and health factors like compulsive use, obesity, sleeping habits, etc. (Dhir et al., 2016; Punamäki et al., 2007). The literature often neglects psychosocial competencies such as emotional self-regulation, risk awareness, and ethical decision-making in digital environments (Jones and Mitchell, 2016; Laaber et al., 2023). Moreover, common measures of digital engagement, including screen time and self-reported frequency of use, fail to capture the detailed behavioral patterns that constitute meaningful digital interactions (Livingstone et al., 2011; Tsetsi and Rains, 2017).

To address these gaps, the concept of digital maturity has emerged as a comprehensive lens to understand how adolescents use technology in ways that support healthy development (Christensen et al., 2025; Laaber et al., 2023; Arenas and Yazdi, 2022). Digital maturity extends beyond technical skills to incorporate a collection of competencies/dimensions including autonomy in navigating digital spaces, respectful digital citizenship, emotional resilience, risk-awareness, and proactive support-seeking that together determine whether an adolescent's digital engagement promotes or hinders well-being (Laaber et al., 2023). Deficits in any area can limit a youth's capacity to benefit from digital tools. However, while the construct of digital maturity offers valuable theoretical insights, its practical utility depends on our ability to translate these dimensions into reliable, data-driven indicators that can inform real-time decisions across educational, community, and digital platform contexts. Such indicators form the backbone of measurement systems that, from an operations perspective, must be precise, scalable, and actionable, enabling stakeholders to allocate available resources whether in schools, community programs, or digital service platforms where it will achieve the greatest impact. Our study responds to this challenge by proposing a data-driven, interpretable approach to assess digital maturity across adolescent populations. We address two central research questions: 1) *Are there distinct, identifiable patterns in adolescents' ICT usage that reflect varying levels of digital maturity?* and (2) *How effectively can adolescents' digital maturity be predicted using interpretable machine learning and to what extent can accuracy be maintained when optimizing for minimal feature acquisition cost?*

We address these by leveraging a large survey dataset from the EU Funded Digymatex project (N $\approx$ 1,440 across Austria, Denmark, Germany, Greece). Our initial analysis on this data found children cluster into three maturity groups of Low, Medium and High digital maturity. We operationalize this through an interpretable machine learning method designed to classify digital maturity levels from ICT usage patterns. Using this dataset, we train an ensemble classifier combining a random forest with a deep neural network to predict these maturity levels from the digital maturity dimensions. Ensembles are selected for their ability to aggregate diverse learners and thus improve generalization accuracy (Dietterich, 2000). This ensemble approach is chosen to balance predictive power and interpretability as the deep neural network captures complex nonlinear relationships among the behavioral indicators, while the random forest provides stability and clearer insights into feature importance. To ensure the model's predictions are transparent and interpretable, we apply Shapley Additive Explanations (SHAP) to attribute predictions to specific behavioral and psychosocial features, allowing decision-makers to pinpoint the operational levers such as improving risk awareness or digital self-regulation that are most likely to shift maturity outcomes.

Finally, to improve the efficiency of data collection and deployment, we introduce a cost-sensitive classification strategy based on Deep Reinforcement Learning (DRL). Instead of administering the full survey to every individual, each time, our DRL agent learns a classification policy that relies only on a small, preselected subset of survey items(features/dimensions) identified as most informative during training. This approach draws on recent advances in 'classification with costly features' paradigm (Janisch et al., 2020), which treats feature acquisition as a sequential decision-making problem. By optimizing the trade-off between predictive accuracy and feature cost, the DRL policy is able to achieve nearly the same accuracy as the full-model classifier while using only a fraction of the input features. In practice, the trained DRL model can identify which few measurements yield the most information about digital maturity, thereby minimizing assessment time and burden. This makes the solution more practical for resource-constrained settings like schools and youth programs, without sacrificing prediction quality. Operationally, limiting inputs to the top three Digital Maturity Inventory (DIMI) items enables brief homeroom or counseling check-ins and supports higher screening cadence (e.g., periodic micro-screens). Because the distilled policy executes using only locally observed items, it aligns with data-minimization principles (GDPR Article 5, 2016) and can be embedded in low-spec devices (Sze et al., 2017), lowering coordination and training overhead. Coupled with a tiered workflow (full survey only for flagged cases), this concentrates scarce counseling time where it has the highest marginal return while preserving decision quality.

Our contributions are threefold. First, we translate the concept of digital maturity rooted in developmental psychology into a compact, interpretable measurement and decision system that can be directly used in schools and youth-facing platforms. By combining the Digital Maturity Inventory (DIMI) (Laaber et al., 2023) with model-based explanations, we identify the specific competencies such as risk awareness, autonomy, and impulse regulation that most strongly influence maturity classification. This moves the idea of digital maturity beyond theory into a form that is measurable, actionable, and sensitive to local context, allowing interventions to target what matters most rather than relying on broad 'screen-time' guidelines. Second, we show how combining ensemble machine learning with a cost-aware deep reinforcement learning (DRL) policy can achieve the same or better accuracy as a full-feature model while using only the top features. We evaluate classification under two regimes: a full-information setting that uses all survey items and a budgeted setting that limits assessment to 3 items per student. A distilled DRL policy, trained with teacher signals from the full model but using only the top-3 items at deployment, recovers full-information accuracy while reducing the number of items by two-thirds. This represents a clear operational gain that the DRL approach finds a Pareto-improving point on the accuracy–cost frontier, demonstrating how intelligent feature selection and knowledge distillation can reduce data collection costs without losing predictive quality. Third, we embed explainability throughout the process. Using SHAP-based feature attribution, we ensure that predictions are transparent and that stakeholders can see more clearly why an adolescent or group is flagged for support. This not only builds trust in the system but also creates a practical 'targeting logic' for allocating resources in schools, technology platforms, and public programs. Together, these contributions show that interpretability and operational efficiency can go hand-in-hand, providing a scalable and accountable way to monitor and improve digital maturity in diverse contexts.

Our study operationalizes data analytics by mapping limited measurements to high-stakes placement decisions under clear acquisition costs, advancing the operations management agenda of moving from prediction to decisions at the point of action (Feng and Shanthikumar, 2023). Methodologically, our cost-aware screening complements recent works that builds early, accurate predictions to trigger timely interventions for disadvantaged populations as we show how to retain accuracy while shrinking inputs, a design that is crucial when measurement itself is the bottleneck (Wang et al., 2024). For practitioners, the takeaway is prescriptive i.e. implement budgeted, interpretable screens to triage scarce counseling and safety resources, monitor competency dashboards, and schedule periodic full-information refresh cycles, a tool for reliable, low-burden decisions in schools and other youth-facing services (Feng and Shanthikumar, 2023; Wang et al., 2024).

1.2 Background

1.2.1 Benefits and Risks of Children's Digital Engagement

Today, children are 'growing up digital' with technology embedded into nearly all aspects of their lives. When used appropriately, ICTs can yield significant cognitive, educational, and social benefits for young people (Livingstone, 2012). Educationally, digital tools greatly expand access to information and individualized learning. Technology and internet *allows knowledge to be accessed more easily*, and that digital technologies can provide personalized learning experiences reaching even disadvantaged groups if deployed ethically and equitably (Rose et al., 2017). In practice, this means that a child in a remote or under-resourced setting can, through the internet, tap into vast libraries of educational content and adaptive learning software that adjusts to their level. Socially and emotionally, digital connectivity can be a lifeline for youth, especially for those who feel isolated or are differently abled (Bakola et al., 2019). Online platforms enable children and adolescents to maintain friendships across distances and forge new social connections beyond their immediate community (Maeneja et al., 2025).

However, alongside the numerous benefits ICTs offer, their problematic usage poses significant challenges that warrant careful consideration, particularly concerning its impact on children, adolescents and young adults. The study by Andreassen et al. (2017) highlights that individuals with addictive social media use tendencies are more likely to be young, have lower levels of education, and belong to lower socioeconomic strata. Moreover, the pervasive nature of ICT exposes children to the risk of cyberbullying and online harassment which has far-reaching consequences. Hinduja and Patchin (2010) identify cyberbullying as a prevalent issue in the digital age, with victims experiencing heightened levels of anxiety, depression, and social isolation. Additionally, studies by Tokunaga (2010) and Patchin and Hinduja (2017) emphasize the role of anonymity and dis-inhibition in online environments, facilitating hostile behaviors that undermine children's emotional resilience and maturity. The widespread use of ICT devices and platforms raises significant concerns regarding children's privacy and security online where Livingstone et al. (2011) emphasize the need for greater awareness among children and parents about the risks associated with sharing personal information online, including identity theft, cyber-stalking, and exposure to inappropriate content. Furthermore, beyond behavioral addictions and online threats, problematic ICT usage extends its broader implications for children's socio-emotional development. Gazzaley and Rosen (2016) highlight the negative impact of excessive screen time on children's interpersonal skills, empathy, and emotional regulation, contributing to a decline in face-to-face interactions and meaningful relationships. Hence, while digital engagements with ICTs is now a near-universal aspect of a young people's life, it comes with significant perils. Recent research also reveals that the loss of face-to-face peer interactions substantially reduces the benefits of online education (Han et al., 2025). The challenge for researchers, parents, and policymakers is to find strategies that maximize the benefits of children's ICT use while mitigating the risks. This delicate balance highlights the need for a deeper understanding of how children use technology and not simply how often, leading to the concept of understanding digital maturity in children.

1.2.2 Conceptualization of Digital Maturity

In developmental psychology and education research, psychosocial maturity, defined as the ability to regulate emotions, make autonomous decisions, and maintain healthy interpersonal relationships, refers to an individual's ability to self-regulate, make responsible decisions, and maintain healthy relationships as they grow into adulthood (Steinberg and Cauffman, 1996; Greenberger et al., 1975). Greenberger et al. (1975) originally proposed a multidimensional model of psychosocial maturity, highlighting three capacities: *individual adequacy* (functioning effectively on one's own), *interpersonal adequacy* (adequate interactions with others), and *social adequacy* (contributions to social cohesion). Steinberg and Cauffman (1996) further refined this concept, focusing on temperance (impulse control), perspective (empathy and consideration of diverse viewpoints), and responsibility (autonomous functioning). Although the lens of psychosocial maturity has been used to understand diverse kinds of behaviors including antisocial behavior, risk-taking behavior, communication skills, etc. (Arenas and Yazdi, 2022), studies analyzing psychosocial maturity and digital technologies are scarce. This work attempts to conceptualize psychosocial maturity in a digital context by capturing behavioral patterns based on children's interaction with digital technologies.

Recent work by Laaber et al. (2023) formally defined digital maturity as *the self-determined, healthy, and adaptive use of digital technology* and developed a comprehensive measurement tool called the Digital Maturity Inventory (DIMI). Rather than treating all screen time as equal, the digital maturity concept recognizes that some children develop the skills, self-regulation, and critical awareness to use ICT in beneficial ways, whereas others struggle to manage their digital lives. In essence, digital maturity describes *'how well' a young person uses technology, not just 'how much'*. It is a multidimensional construct (DIMI) that spans technical competencies, behavioral patterns, and psychosocial factors. The DIMI is composed of ten dimensions that characterize a digitally mature child. These dimensions include both capabilities and habits, as described in Table 1.1.

Each dimension contributes to the holistic profile of a child's digital maturity. A youth might excel in certain areas but lag in others. For instance, a tech-savvy teen with high digital literacy might still be prone to impulsive scrolling at night indicating low autonomy or regulation. The interplay of all these competencies determines whether a child's overall relationship with technology is healthy, that is adaptive and enriching, or problematic, which may be risky or detrimental. Empirical studies lend support to digital maturity as a meaningful measure (Christensen et al., 2025; Hofmans et al., 2024; Laaber et al., 2025; Koch et al., 2024). A qualitative grounded-theory study by Arenas and Yazdi (2022) gathered insights from focus groups with children, generating 'maturity factors' that closely parallel the dimensions above (e.g., technology awareness, perceived benefits vs. risks, usage patterns, emotional self-regulation, and digital civic behavior). This convergence suggests that both experts and youths perceive a similar set of attributes defining mature digital use.

1.2.3 Challenges in Assessing Digital Maturity with Traditional Approaches

Measuring a child's digital maturity is inherently challenging, given the construct's multifaceted nature. Early research in this field often relied on simplified metrics that only partially capture the concept, such as checklists of digital skills or crude indicators like total screen time per day (Radesky et al., 2016). Although these methods provided initial insights, scholars have increasingly recognized their limitations. These measures can quickly become outdated as technology evolves. Additionally, there has been a reliance on single-dimension measures, such as total screen time (Ponti, 2023). The number of hours spent online or devices owned is relatively easy to quantify and thus frequently employed as proxies for digital engagement. However, such metrics fail to distinguish between the content or quality of digital interactions (Orben and Matias, 2025; Laaber et al., 2023; Christensen et al., 2025). For example, three hours spent coding or reading e-books vastly differ from three hours spent scrolling through videos or playing violent games. Previous research has often narrowly focused on academic performance or time limits, lacking comprehensiveness and neglecting broader psychosocial profiles. In practice, a child's well-being might depend less

Table 1.1: Digital Maturity Dimensions and Working Definitions

Digital Maturity Dimension	**Definition**
Digital Literacy	Knowing how to find information online, judge credibility, and protect privacy and data. A digitally literate child can navigate the internet safely and intelligently.
Autonomy of Choice	Using devices by deliberate choice and purpose rather than compulsion. A mature user can self-regulate screen time and avoid impulsive checking.
Individual Growth	Leveraging technology for personal development and learning, for example using educational apps or creative software to build skills rather than only consuming content.
Risk Awareness	Recognizing online dangers such as scams, contact risks, and misinformation, then making prudent choices to avoid harm. A mature child applies basic cyber-safety practices.
Support-Seeking Behaviour	Willingness to ask for help when facing digital problems or distress. Instead of hiding issues like cyberbullying, the child turns to a parent, teacher, or peer.
Regulation of Aggressive Impulses	Managing urges toward hostility in digital settings. A mature user restrains negative reactions and opts for constructive communication over retaliatory posts.
Regulation of Negative Emotions	Recognizing, understanding, and controlling adverse emotions triggered online. Digitally mature children manage frustration, anxiety, or envy and maintain emotional stability.
Autonomy within Digital Context	Independently navigating and managing digital interactions and environments. A mature child sets personal boundaries, engages purposefully, and keeps control without constant external guidance.
Respect Towards Others	Showing empathy and courtesy in digital exchanges. A mature user follows netiquette, avoids harassment, and understands the impact of words and actions online.
Digital Citizenship	Acting responsibly and positively in online communities, from reporting harmful content to participating in civic or volunteer projects via digital platforms.

on the amount of time spent online and more on the nature and emotional impact of their online activities; factors that duration metrics alone fail to capture.

Recognizing these gaps, researchers and child advocates have advocated for more comprehensive measurement tools (Orben and Matias, 2025). Initiatives such as the International Computer and Information Literacy Study (ICILS) (Fraillon, 2024) and the EU Kids Online project (Smahel et al., 2020) collect data not only on the frequency of online activities but also on the experiences, both positive and negative, that children encounter in digital environments. Yet, even these advanced initiatives face inherent constraints as they rely on periodic data analysis methods. To address the challenges identified, recent research has utilized machine learning (ML) and predictive analytics (Arpaci, 2023) applied to children's ICT usage data, providing a novel means of assessing digital maturity. This approach capitalizes on data analyzed within robust frameworks to infer patterns indicative of healthy or risky digital engagement. This aligns with a broader trend in educational and behavioral sciences, employing data (Baker et al., 2016) analytics for personalized assessments and interventions.

We address these challenges by replacing single-exposure proxies with an interpretable, low-burden decision pipeline that schools can run as part of routine operations. We (i) derive maturity profiles from DIMI dimensions using a Gaussian Mixture Model and (ii) train a transparent RF+DNN classifier with SHAP attribution, then (iii) deploy a distilled DRL policy that assigns profiles using only the top three items at the point of use. For education operations, this design enables brief homeroom or counseling micro-screens, tiered support (full survey only for flagged cases), and competency dashboards that guide scheduling and staffing to where need and payoff are highest. Because the policy uses few items and provides feature-level explanations, it reduces measurement burden, supports accountability and communication with parents and staff, and fits within capacity-constrained settings while maintaining accuracy—consistent with recent POMS guidance on operational data analytics and early, action-oriented prediction in service systems (Feng and Shanthikumar, 2023; Wang et al., 2024).

1.3 Data and Methodology

We approach the assessment of children's digital maturity as a two–part problem that links understanding of digital maturity to its accurate classification. First, we uncover and then classify meaningful maturity profiles from observed behavior. Second, we deliver those classifications with as little measurement as possible while preserving accuracy for routine use in schools. Because the underlying relationships among maturity dimensions are nonlinear and potentially high–dimensional, we rely on machine learning methods that are well suited to capture such patterns (Benbya et al., 2020). Figure 1.1 summarizes the overall methodology.

Part I: Understanding and Classification. We identify latent maturity groups using a Gaussian Mixture Model (GMM). Unsupervised learning is appropriate here because there are no 'ground truth' labels for digital maturity, and GMM provides soft memberships and BIC–based model selection that respect fuzzy developmental boundaries (Zhang and Chen, 2019; Shreem et al., 2022; Hockenberry et al., 2011). We fit the GMM on the training split and apply the learned mixture to the hold-out set to prevent information leakage. Using these cluster–derived labels as targets, we then train a supervised classifier to predict an individual's maturity level from the full set of post-processed features. For this, we choose a heterogeneous ensemble that combines a Random Forest (RF) with a modest Deep Neural Network (DNN) where RF offers stability and interpretable feature signals (Breiman, 2001), while the DNN captures residual nonlinearities typical of behavioral and psychosocial data (Zhang et al., 2024). Heterogeneous ensembles frequently improve generalization by aggregating diverse inductive biases (Dietterich, 2000; Sherazi et al., 2021). The resulting RF+DNN model, evaluated with weighted soft voting, is our *full-information* benchmark model because it uses all available features. Post-hoc explanations translate predictions back into the language of dimensions for interpretability.

Part II: Cost–Aware Classification with DRL. We then address the operational question: can we retain performance while drastically reducing inputs so that brief screens are feasible

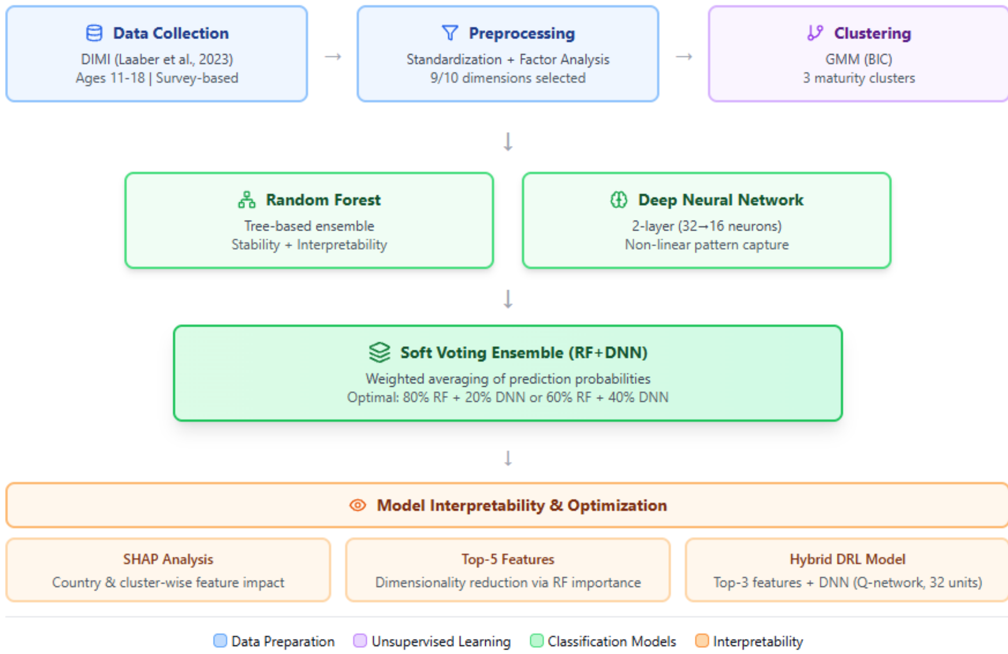

Figure 1.1: Research Methodology

in practice? We frame the task as 'Classification with Costly Features' (Janisch et al., 2020). The objective is to maintain predictive quality while limiting the number of survey items per student. Building on evidence that reinforcement learning can outperform static rules in difficult classification settings (Lin et al., 2020), we train a Deep Reinforcement Learning (DRL) policy that produces the maturity label in a single decision step (De Asis et al., 2018). The policy's *state* is constructed from a small set of country-specific Top-3 dimensions identified on the training data *plus* an *expert signal* from a full-information teacher model (the full information benchmark) during training. This teacher guidance distills global patterns into the policy. At deployment the policy observes only the Top-3 items and does not query the teacher. In short, training uses Top-3 + teacher probabilities to learn, while the final model uses only the three inputs to predict, yielding a low-burden classifier aligned with school constraints without sacrificing accuracy.

1.3.1 Data Preparation

For our analysis, we utilize data from the Digital Maturity Inventory (DIMI), a validated survey capturing diverse psychosocial and behavioral dimensions of adolescents' ICT usage (Laaber et al., 2023). This dataset was selected as it was originally collected to measure children's digital maturity and due to its comprehensive coverage of digital maturity constructs and international scope, allowing for robust cross-country comparisons. Our sample comprises adolescents aged 11–18 across four European countries: Austria (n=603), Germany (n=406), Greece (n=284), and Denmark (n=147), totaling 1,440 valid responses. Detailed demographics are presented in Table 1.2.

The items measuring *Digital Literacy* assessed the children's knowledge of privacy set-

Table 1.2: Descriptive statistics by country

	Austria		**Germany**		**Greece**		**Denmark**	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Initial sample size	603		406		284		147	
After outlier removal	556		378		251		134	
Number of observations	293	263	185	193	133	115	67	67
Average age (years)	15.11	15.17	15.24	15.21	15.54	15.72	15.12	15.59
SD (years)	1.85	1.93	1.95	2.00	2.13	2.15	1.75	1.74

tings on social media sites, ability to disable website cookies and understanding of how to store files on the cloud. *Risk Awareness* was assessed by asking children about their level of caution when using digital technologies and the importance they placed on their own safety while online. *Individual Growth* was indicated by the extent to which children were learning new skills and useful information through their use of digital technologies. *Respect Towards Others* was assessed by considering the children's attitudes toward others, their ability to respect the opinions of others while online, and their use of appropriate language when disagreeing with others. *Digital Citizenship Behavior* was assessed through items that explored the extent to which children used technology to improve their local communities, support environmental campaigns, and stand up for important issues. *Support Seeking Behavior* measured the extent to which children sought help from parents, siblings, or friends when encountering technical or social problems online. *Autonomy Within Digital Context* and *Autonomy of Choice* dimensions collectively measured impulsive behavior among children. Autonomy Within Digital Context assessed the children's ability to choose what they did while using digital devices, while Autonomy of Choice explored the children's autonomy in choosing to be online or not, including factors such as fear of missing out. *Regulation of Aggressive Impulses* assessed children's reactions to criticism or insults received online and whether they acted impulsively without considering the consequences or regretted their actions later. Finally, *Regulation of Negative Emotions* measured the extent to which children were affected by negative experiences online, such as becoming upset or annoyed, and how long it took them to recover from these experiences.

To ensure high data quality, we perform a series data cleaning steps before we proceed downstream in our analysis. First, we performed factor analysis to confirm the DIMI's theoretical structure initially comprising ten dimensions as mentioned in Section 1.2.2. Employing Varimax rotation, an orthogonal method that maximizes the interpretability of factors (Allen, 2017; Dormann et al., 2013), we aimed to achieve approximately orthogonal dimensions to avoid multicollinearity issues in our predictive models. The analysis showed redundancy within dimensions across countries, prompting the removal of one redundant dimension per country to maintain nine distinct, orthogonal dimensions. In Austria the dimensions *Respect Toward Others* and *Regulation of Aggressive Impulses* collapsed onto one factor; in Germany, *Regulation of Negative Emotions* and *Autonomy of Choice* merged; similar pairings occurred in Greece and Denmark. In each case, we dropped one of the two highly correlated dimensions (choosing the one whose removal yielded a cleaner factor structure) to enhance the interpretability and stability of the model and also to streamlines the practical implementation of our assessment tool, reducing respondent fatigue and operational complexity (with fewer inputs).

Furthermore, as clustering algorithms are sensitive to noise and extreme values, we next applied the Isolation Forest anomaly detection algorithm (Liu et al., 2008) to identify and remove outlier responses in each country's dataset. Isolation Forest is an ensemble method that isolates anomalies by randomly partitioning data, making it effective in detecting aberrant observations that could distort the shape and boundaries of clusters. Removing these anomalies reduces noise and improves the reliability of the unsupervised grouping step, ensuring that cluster definitions reflect genuine patterns rather than being skewed by atypical responses. By the end of this preprocessing, we had a refined set of 9 features for each country's dataset, each representing a distinct aspect of digital maturity. This prepared feature set would serve as the input for clustering and modeling steps.

1.3.2 Determining Digital Maturity Levels

In the absence of predefined labels for digital maturity, we employed unsupervised machine learning to uncover clusters of children with similar maturity profiles. Clustering provides a data-driven way to define categorical maturity levels. We experimented with standard clustering algorithms including k-means and k-modes for categorical variants, but ultimately selected a Gaussian Mixture Model (GMM) clustering approach. GMM was chosen for several reasons: *(i)* it assumes the data is generated from a mixture of Gaussian distributions, which is suitable for continuous Likert-scale survey data; *(ii)* it provides 'soft' cluster assignments with membership probabilities, acknowledging that some children might lie between categories (Murphy, 2012); and *(iii)* it offers a principled way to determine the optimal number of clusters via model selection criteria like the Bayesian Information Criterion (BIC). Soft clustering can capture overlapping group boundaries more effectively than hard partitioning methods. We used the BIC to evaluate cluster solutions with different numbers of clusters for each country's data. The BIC balances model fit with complexity, penalizing excessive clusters, and is commonly used to choose the number of mixture components.

Based on BIC criteria (Figure A1) we found that a three-cluster solution was optimal for Austria, Greece, and Denmark, whereas for Germany the lowest BIC was achieved with two clusters with the three-cluster solution being a very close second. The difference in BIC for Germany's 2 vs. 3 clusters was marginal (ΔBIC= 3), suggesting that a third cluster would not degrade the model much. To maintain consistency and interpretability across countries, we decided to adopt three clusters for all four countries. This decision allows us to define analogous maturity levels. Adopting a common three-level categorization for all countries facilitated cross-country comparisons and a richer analysis of group differences. We fit the GMM clustering on the 70% training split and then assigned cluster labels to the 30% hold-out for evaluation, for the 9 dimensions for each country after the pre-processing steps. Each child in the dataset was thus assigned to one of three clusters. By using clustering in this way, we have a proxy ground truth for digital maturity categories. These cluster-derived labels for each child became the target variable for the next phase of our study which was the supervised classification modeling. This approach of generating labels via unsupervised learning, then training a classifier, is an established strategy when true labels are unavailable, as it combines the strengths of discovery and prediction (Hockenberry et al., 2011).

1.3.3 Supervised Ensemble Classifier

With each child now labeled by a maturity cluster, we turned to supervised learning to build a model that can predict a child's maturity level from their responses. Our goals for this classifier were twofold: high accuracy and interpretability. We chose an ensemble classifier combining a Random Forest (RF) and a Deep Neural Network (DNN). The rationale is that these two models offer complementary strengths. Random Forests, an ensemble of decision trees, are robust and relatively interpretable; they handle nonlinearity and interactions and provide measures of feature importance. DNNs capture complex, higher-order relationships that tree-based models might miss (Breiman, 2001; Zhang et al., 2024). Concretely, the ensemble learns a mapping, $f : \mathbb{R}^d \to \mathcal{Y}$ from the d post-processed DIMI dimensions (here d=9) to maturity classes, where $\mathcal{Y}$ is the finite set of maturity classes discovered by the GMM ($|\mathcal{Y}| = 3$), and produces calibrated probabilities $p(y \mid x)$. The RF contributes threshold-like rules and low-order interactions among dimensions, while the DNN contributes smooth, higher-order combinations that reflect psychosocial synergies. We do this also because these probabilities and their attributions translate directly into actions, *i)* providing the full-information benchmark used later for our cost-aware policy, *ii)* indicating which individuals are likely to belong to each maturity class, and *iii)* revealing which competencies (e.g., risk awareness, autonomy within context, impulse regulation), thereby supporting targeted supports. By combining a symbolic learner (RF) with a sub-symbolic learner (DNN), the ensemble strengthens the diversity of learning biases. Moreover, ensemble methods are known to improve generalization by aggregating multiple classifiers' outputs (Dietterich, 2000). Prior work shows heterogeneous ensembles can outperform single models across domains; for example, Sherazi et al. (2021) report that hybrid ensembles with neural networks achieve superior accuracy in medical diagnosis. We expected that in our context, the RF might excel in cases where simple combinations of features indicate maturity, whereas the DNN might detect subtler patterns or combinations of behaviors that do not have well-defined thresholds. Moreover, to ensure there is no leak of information from clustering into the generated labels for the predictive classifiers, a holdout dataset was kept aside from the clustering and training of the predictive classifiers and was used only for testing purposes. This ensures that there is no overfitting in our classifiers and that the accuracies are not over-optimistic. 30% of the datasets were used as the holdout dataset.

Random Forest: We implemented a Random Forest classifier (Breiman, 2001) as one component of the ensemble. The RF was chosen for its robustness against overfitting (due to averaging many decision trees) and its interpretability in terms of feature importance. We performed grid search cross-validation on key RF hyperparameters (number of trees, maximum tree depth, minimum samples per split) for each country's training set to ensure the model was well-tuned to the data. Each RF produces, for a given input, a probability distribution across the three clusters, obtained by the fraction of decision trees voting for each class. This probabilistic output of RF was important for combining with the DNN. We also saved the RF's feature importance rankings (Gini importance) for later use in feature selection and interpretation.

Deep Neural Network: In parallel, we developed a feed-forward neural network to serve as the other component of the ensemble. Given the relatively small dataset (each country had only a few hundred children after filtering), we kept the network architecture modest to avoid overfitting. After experimentation, we settled on a fully-connected DNN with two hidden layers of 32 and 16 neurons respectively with ReLU activation (Goodfellow et al., 2016; LeCun et al., 2015). The output layer is a softmax layer yielding class probabilities for the three clusters. We applied L2 weight regularization and a dropout rate of 0.2 to the hidden layers to further guard against overfitting, following best practices for training on small data (Goodfellow et al., 2016; Nair and Hinton, 2010; Glorot et al., 2011). The network was trained using the Adam optimizer (Kingma and Ba, 2014) with an initial learning rate of 0.01. The DNN's architecture and hyperparameters (layer sizes, regularization strength, learning rate) were chosen via grid search to balance bias and variance. By including a neural network alongside the RF, we allow the overall model to potentially detect subtle nonlinear patterns or interactions among the maturity dimensions that a decision tree might not capture easily.

Soft Voting Ensemble: To combine the RF and DNN, we used a weighted soft voting ensemble strategy. In soft voting, each model produces a probability for each class, and we average these probabilities to make the final prediction. For example, if the RF predicts a probability vector ($p_{\text{RF},1}, p_{\text{RF},2}, p_{\text{RF},3}$) and the DNN predicts ($p_{\text{DNN},1}, p_{\text{DNN},2}, p_{\text{DNN},3}$), and w is the weight for each strategy then the ensemble prediction is based on p_{ensemble} =w.p_{RF}+(1-w).p_{DNN} with $w \in 0.8, 0.6, 1.0$ for each cluster, and we choose the cluster with the highest averaged probability. This approach takes into account each model's confidence in each class, unlike hard voting where each model would simply cast one vote for its top class. Soft voting often yields better performance because it uses more information from each model's output, and it effectively gives a weighted influence to models' predictions proportional to their confidence. In our case, we evaluated three configurations: (1) RF80–DNN20, where the final prediction was based on 80% RF probability and 20% DNN probability; (2) RF60–DNN40; and (3) standalone models using 100% RF or 100% DNN. The final RF+DNN ensemble was trained on each country's data using the clustered labels as targets respectively. We used a 70/30 train-test split for each country, and within the training set we performed 5-fold cross-validation during hyperparameter tuning. The ensemble's performance on each country's test set represents the baseline accuracy achievable when using the full set of 9 orthogonal features.

Feature Reduction: Top-5 Feature Model

A key question for operational efficiency is whether we can achieve similar predictive accuracy using fewer input features. Shorter questionnaires are highly desirable in practice to reduce respondent fatigue and increase adoption, as long as predictive power is not severely compromised. To investigate this, we conducted a feature reduction experiment after training the full models. Using the trained Random Forest on the full 9 features, we extracted its feature importance rankings. We also cross-checked these with feature importance derived from SHAP values (SHapley Additive exPlanations) for the ensemble, to ensure we picked truly influential features and we found there was substantial agreement between RF Gini importance and SHAP importance on top features. We then selected the top 5 features that were most influential for the RF's predictions in each country. These five dimensions can be seen as a condensed subset of the survey that carries most of the information needed for classification. Using only these top 5 features, we retrained the same ensemble classifier (RF + DNN with the same hyperparameters) for each country. We refer to this as the Top-5 feature model. The training procedure was the same, except the input vector now had 5 dimensions instead of 9. This allowed us to directly compare the performance of a model using all features versus using nearly half the features, to quantify the trade-off between model simplicity and accuracy. Prior research in machine learning suggests that judicious feature selection can sometimes maintain (or even improve) accuracy by eliminating noisy or redundant features. However, there is often a drop in performance if truly informative predictors are removed. If the accuracy drop was modest, it would validate the idea that a much shorter questionnaire (5 key questions instead of the full set) could suffice for classification. On the other hand, a large drop in accuracy would highlight that the remaining features (ranks 6–9) still carried important information. This feature reduction test sets a baseline for how well we can do with minimal inputs without any advanced techniques.

1.3.4 Hybrid Deep Reinforcement Learning Classifier

To attain high accuracy with minimal features, we introduced a novel hybrid Deep Reinforcement Learning (DRL) classifier. The DRL policy is trained with 'teacher distillation' from a full-information DNN, while its state is restricted to the country-specific Top-3 raw items. Unlike prior costly full-feature RL approaches that acquire features sequentially at inference, our policy deploys as a single-step classifier that observes only those three items and aims to recover full-information accuracy. The motivation behind this approach was to transfer knowledge from the complex full-feature model into a simpler model that uses only a few key features, aiming for high accuracy with low-dimensional input. In machine learning, such transfer learning (Zhu et al., 2017, 2023), that is, knowledge gained from a rich, high-dimensional task can be exploited to improve learning in a related task with fewer inputs. Within operations research, the *'classification with costly features'* paradigm addresses the trade-off between predictive accuracy and data-collection cost by modeling feature acquisition as sequential decision making (Janisch et al., 2020; LeCun et al., 2015; Suttle et al., 2021). DRL agents can learn cost-accuracy trade-offs that minimize survey length while preserving performance. We formulated the maturity classification problem as a single-step decision in a reinforcement learning framework (Markov Decision Process) to apply DRL. Essentially, we treat the prediction of the maturity cluster as the action an RL agent must take after observing a state that contains only limited information about the child.

State Representation: To represent the stated, we concatenate a small number of most predictive raw features with an informed hint from the full model's prediction. Specifically, the state vector included the top 3 raw features (from the RF importance ranking) for a given child, along with the class-probability vector from the full 9-feature DNN model for that child. The DNN's output probabilities serve as an expert recommendation or summary of the child's profile in the context of the full feature space. By incorporating these into the state, we give the DRL agent direct access to a distilled form of the complex patterns that the full model has learned. This is a form of knowledge distillation where the DRL model uses the full DNN's prediction as an additional input, using the high-dimensional knowledge without having to process all those features itself. Importantly, these DNN-derived probabilities are used only during training (or occasional model refresh cycles) to transfer knowledge; during normal deployment, the DRL policy operates using only the Top-3 raw features. Formally, for each child we define the state as:

$$s = [x_{(i_1)}, x_{(i_2)}, x_{(i_3)}, P_{\text{DNN}}(Y=1), P_{\text{DNN}}(Y=2), P_{\text{DNN}}(Y=3)],$$

where $x_{(i_1)}, x_{(i_2)}, x_{(i_3)}$ are that child's values on the top-3 features, and $P_{\text{DNN}}(Y)$ are the predicted probabilities from the full-feature DNN. This yields a 6-dimensional state vector and as compared with the original 9-dimensional feature space, now the DRL's state is much more compact, yet being augmented with high-level knowledge from the full model. This hybrid state design draws inspiration from transfer learning approaches in RL where an agent is provided with expert advice or demonstration to accelerate learning (Zhu et al., 2023). In our case, the full DNN's prediction acts as an 'expert' guiding the RL agent in the reduced state space.

Actions and Reward: The DRL agent's action a is to select one of the three maturity labels for the child. This is a single-step decision problem where the agent observes the state s (the child's top-3 features + DNN recommendation) and must immediately take an action. After the action, the environment provides a reward r and the episode ends. We designed the reward signal to strongly encourage correct classification. The agent receives a large positive reward for a correct prediction and a negative reward for an incorrect prediction.

Formally, we define

$$r \;=\; \begin{cases} R_+ & \text{if the predicted label matches the true label,} \\ R_-^{(1)} & \text{if the prediction is off by one class (adjacent),} \\ R_-^{(2)} & \text{if the prediction is off by two classes,} \end{cases}$$

with $R_+ > 0$, $R_-^{(2)} < R_-^{(1)} < 0$. This reflects the ordinal structure of the classes: the worst mistakes (two levels away) are penalized more than adjacent mistakes. In the interest of calibration and stability, we tuned the *ratios*

$$\gamma_1 = \frac{|R_-^{(1)}|}{R_+}, \qquad \gamma_2 = \frac{|R_-^{(2)}|}{R_+}$$

on the training split and selected $(\gamma_1, \gamma_2) = (0.25,\ 0.50)$, which balances a strong incentive for correct decisions with graded penalties for near vs. far errors. For reporting clarity we set $R_+ = 1$, $R_-^{(1)} = -0.25$, $R_-^{(2)} = -0.50$ (for example, rewarding/penalizing $+200/-50/-100$ respectively). Because each episode ends after one step, multiplying all rewards by a positive constant leaves the optimal action unchanged; what matters are the relative magnitudes (Suttle et al., 2021). We kept the reward class-agnostic (no per-class weights) and evaluated fairness via macro-F1, which gives each class equal importance at test time. This shaping focuses learning on achieving high accuracy while preserving a meaningful margin between 'near' and 'far' mistakes.

Learning Algorithm: We implemented the DRL classifier using a Deep Q-Learning (DQN) approach (Mnih et al., 2015). In this setup, the agent uses a neural network to approximate the Q-value for each state-action pair, *Q(s, a)*, which represents the expected reward of taking action a in state s. Since each episode ends after one action, the Q-value of a state-action is essentially just the expected immediate reward for that classification. The DRL agent's policy during training was ϵ-greedy that is, it chooses a random action with probability ϵ (to explore) and the best estimated action with probability $1-\epsilon$. We varied the value of ϵ over the course of training to ensure ample exploration in early stages and more exploitation later. The Q-network architecture was a simple multilayer perceptron taking the 6-dimensional state as input, with one hidden layer of 32 neurons and an output layer of 3 neurons (each corresponding to the Q-value of one action). We chose this architecture for its computational efficiency and to mitigate the risk of overfitting, which is a common challenge with more complex networks on small-dimensional data. The policy is distilled from the full DNN during training; at deployment, the agent uses only the three features via a Q-network with one hidden layer of 32 units. This simple design is a well-established baseline for tabular like RL problems (Mnih et al., 2013). We trained this network by sampling training examples and using the DQN update rule after the agent takes an action and receives a reward, we compute the error between the Q-network's prediction *Q(s,a)* and the target value. We also utilized standard DQN techniques such as experience replay (storing past state-action-reward examples and training on random batches of them) and a separate target network to stabilize training (Mnih et al., 2015). We trained the agent for many epochs, cycling through the training set of children repeatedly until the policy converged i.e. the training accuracy plateaued.

Model Evaluation and Comparison: Our primary evaluation metrics were classification accuracy and F1-score. Accuracy is a straightforward measure of overall correctness, while the macro F1-score is the average of the F1-scores for each class and thus treats each class equally. We report both because in our datasets the class distribution is not perfectly balanced. A model could achieve high accuracy by doing well on a majority class, but the macro F1 would highlight if it performs poorly on a minority class. Using macro F1 ensures that performance on each maturity level is given due attention. To compare the models, we focused on three in particular for each country: (1) the Full-Feature Ensemble (RF+DNN on all 9 features), which represents the benchmark with maximal information; (2) the Top-5 Feature Ensemble, which uses a reduced set of 5 features; and (3) the DRL (Top-3 + DNN) model, which uses only 3 raw features plus the DNN's transferred knowledge. To determine if differences in performance were statistically significant, we conducted paired t-tests on the F1-scores from the 30 runs of each model. For each country we compared the DRL vs. full model F1 across the 30 paired observations, as well as DRL vs. top-5, and full vs. top-5 allowing us to test whether for example, the DRL significantly outperform the simple top-5 model.

1.3.5 Model Interpretation with SHAP

To ensure our classifier is interpretable and to extract insights for each cluster, we employed SHAP (Shapley Additive Explanations) for post-hoc explanation of the model's predictions. SHAP is a framework based on cooperative game theory that assigns each feature a contribution value for a particular prediction (Lundberg and Lee, 2017). SHAP calculates how much each feature increases or decreases the predicted probability of a certain class relative to a baseline. It is a well-known techniques for interpretable machine learning due to its robust theoretical foundation and consistent feature attribution (Lundberg and Lee, 2017). We compute SHAP values for all the classifiers for each country in our study allowing us to produce: *(a)* A feature importance rankings for each country and *(b)* class-specific interpretation by focusing on SHAP values for each class label (digital maturity cluster). The use of SHAP enables us to translate the predictive model back into the language of the digital maturity dimensions, thus providing explanations that are meaningful to stakeholders. Rather than a black-box ‘accuracy tool’, the model doubles as an analytic lens to understand why a given child may be classified in any particular cluster.

1.4 Findings

Our primary focus was to identify clusters of children with different digital maturity levels. We identified 3 clusters each for the four countries. Table 1.3 shows the descriptive statistics for each cluster.

Table 1.3: Descriptives by country with clusters in columns

	Group 1			**Group 2**			**Group 3**		
Country	n	Age (SD)	M/F	n	Age (SD)	M/F	n	Age (SD)	M/F
Austria	139	15.81 (1.77)	77/62	245	14.96 (1.92)	121/123	172	15.05 (1.85)	95/77
Germany	109	15.67 (1.91)	51/58	143	15.49 (1.97)	72/71	126	14.54 (1.89)	62/64
Greece	75	15.70 (2.08)	37/36	50	16.03 (1.83)	29/21	126	15.40 (2.24)	68/57
Denmark	37	15.44 (1.81)	18/19	43	15.43 (1.87)	20/23	54	15.36 (1.67)	29/25

To identify unique traits and similarities among the identified clusters, we generated a composite graph in Figure 1.2. This plot compared the digital maturity dimension score of each cluster group against their mean z-score, with each subplot corresponding to one of the four countries. The mean z-score values, representing the deviation from the average value of the entire dataset, were used to evaluate the performance of the cluster groups across different dimensions. A positive z-score indicated that a particular cluster group performed better than the average in that dimension, while a negative z-score suggested poorer performance. This graph provided a visual representation of the relative performance of each cluster group, facilitating the identification of clusters with above or below-average performance.

While the general pattern holds in all countries, there are some notable cross-country differences in cluster characteristics, echoing cultural or contextual factors. Drawing on the original study’s detailed cluster profiling , we summarize each cluster’s profile and how it manifests in the four countries:

- **High Digital Maturity Group:** This group stands out for advanced digital skills and responsible behavior. Children in this cluster score well above average on most dimensions. They tend to use ICTs in constructive ways that support personal growth and healthy social relationships. For example, in Austria, High maturity children showed exceptional Digital Literacy, strong Autonomy within Digital Context, and

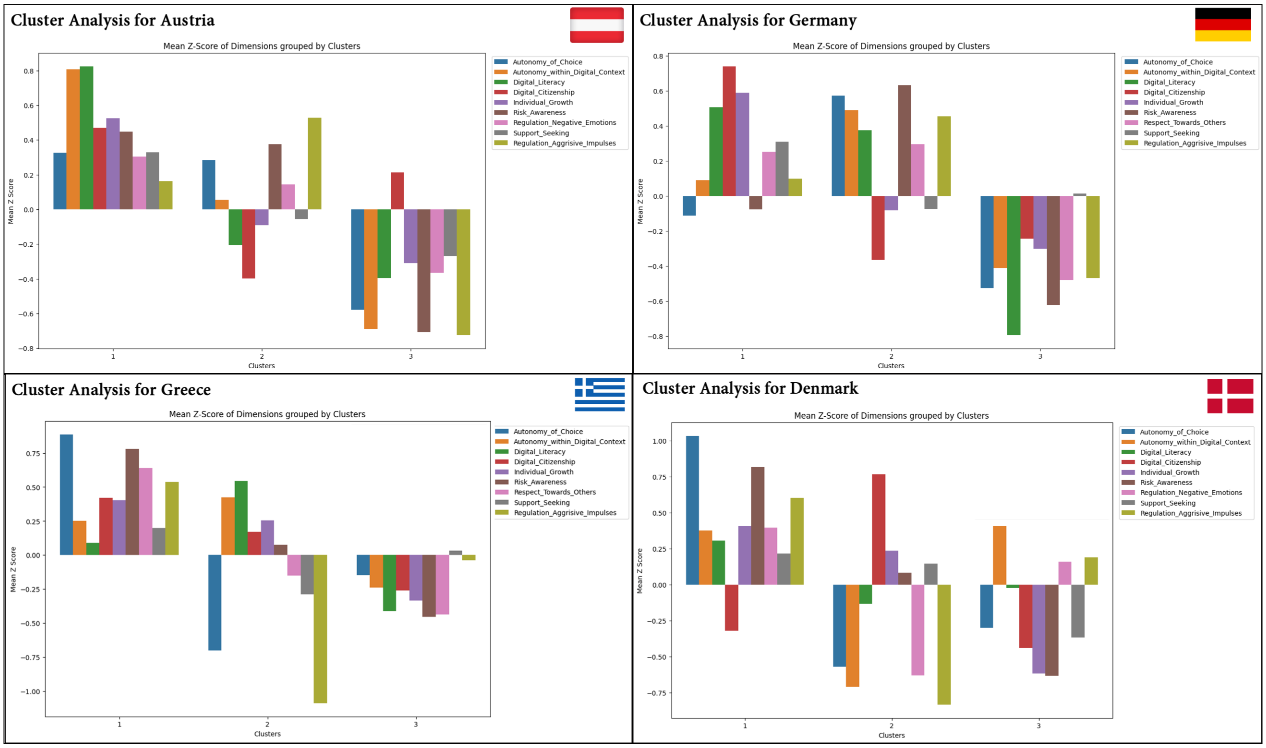

Figure 1.2: Clusters Obtained from Austria, Germany, Greece and Denmark

high Individual Growth, indicating they adeptly learn new skills online and use devices responsibly for self-improvement. They also exhibited non-compulsive usage (low impulsivity). German children in this group excelled in Digital Citizenship and Individual Growth, suggesting engagement in positive online community activities and learning, though interestingly they were only average in Risk Awareness. Greek children in the high group demonstrated remarkable Autonomy and strong Risk Awareness, along with positive social behavior, reflecting a balance of independence and cautiousness online. Danish children in this cluster showed proficiency in autonomy and notably good Emotional Regulation, meaning they handle upsetting online content or conflicts calmly. However, Danish high-maturity kids struggled a bit with Digital Citizenship and risk awareness, potentially indicating that while they are personally skilled and well-regulated, they may not be very involved in online civic matters. Overall, the high maturity cluster in all countries represents well-rounded 'digitally mature' youths as they are indicative of using the internet confidently and positively, have low levels of maladaptive behaviors, and maintain healthy control over their digital lives. Minor weaknesses (like Germans' lower risk awareness or Danes' lower civic participation) point to specific areas where even these advanced users can improve.

- **Medium Digital Maturity Group:** This group (also referred to as Average maturity) shows a mix of above and below average scores across dimensions. They are intermediate in their digital habits, generally responsible in some respects but lacking in others. The medium cluster tends to have certain strengths, often in self-control, but also clear weaknesses in more complex or externally-oriented skills. For instance, Austrian children in this cluster were above average in Autonomy of Choice, Regulation of Aggressive Impulses, and Risk Awareness, indicating they are good at avoiding impulsive online behaviors and are mindful of safety. However, they scored below average in Digital Citizenship, Individual Growth, Digital Literacy, and Support-Seeking. This suggests that while they behave prudently, they engage less in proactive skill-building or community-oriented activities online. In Germany, the medium group similarly had strengths in emotional regulation but weaknesses in Digital Citizenship, Individual Growth, and Support-Seeking. This paints a picture of German medium-maturity kids as generally polite and self-controlled online, yet not deeply utilizing ICT for self-improvement or reaching out for guidance. Greece's medium cluster was somewhat inverted: Greek children here excelled in Digital Literacy and Autonomy within Digital Context but struggled with Regulation of Negative Emotions and Autonomy of Choice, and had low support-seeking behavior. In other words, they are tech-savvy and bold online but may react poorly to upsetting content (e.g. taking criticism badly) and exhibit some compulsive tendencies. In Denmark, the medium cluster showed strengths in Regulation of Aggressive Impulses and Risk Awareness, but all dimensions related to Individual Growth were notably below average. Danish children of medium maturity thus avoid aggressive or risky behavior (a positive sign) yet are not leveraging ICT for learning new skills or personal development. They might use digital media more passively or purely for entertainment rather than growth. Summarizing the medium cluster, these children often have the basics of self-regulation down but they lack initiative or breadth in positive digital engagement. They could benefit from encouragement to use technology more constructively and to improve areas like seeking support or coping with negative experiences.

- **Low Digital Maturity Group:** This cluster comprises the children with the lowest scores on most digital maturity dimensions. They exhibit the most problematic usage patterns like impulsive or uncontrolled ICT use, low digital skills, and poor adaptation. Across Austria, Germany, Greece, and Denmark, the low group showed predominantly negative z-scores on nearly all dimensions. For example, in Austria the only dimension where the low group was not below average was Digital Citizenship, which was slightly above average. This is an interesting anomaly as perhaps some Austrian low-maturity kids express support for causes online even though they lack many other skills. Aside from that, Austrian low group children had very high compulsive behavior and poor scores in literacy, support-seeking, etc., indicating frequent unsupervised use without guidance. In Germany, the low group had one small relative strength, slightly higher-than-average Support-Seeking behavior, but they were very low in Regulation of Negative Emotions, Risk Awareness, and Autonomy within Digital Context. This suggests German low-maturity children often act impulsively and emotionally online, have little awareness of dangers, and cannot navigate digital environments safely on their own. Greece and Denmark showed similar patterns, low-maturity youths were characterized by compulsive usage, poor emotional regulation, and limited digital literacy. In Denmark, almost every dimension was negative. In Germany, Digital Literacy and Risk Awareness were identified as particularly low in this cluster meaning these children may not even know how to use privacy settings or understand basic online safety. Broadly, the low maturity cluster appears to be a high-risk group as these children are more susceptible to negative online experiences and likely exhibit maladaptive behaviors like cyber-aggression or overuse. They lack the self-regulatory skills to restrain unhealthy impulses and the knowledge to keep themselves safe. This highlights the need for targeted interventions for this group to raise their digital competencies and coping strategies.

1.4.1 SHAP Interpretation Across Countries

SHAP analysis provides interpretability for machine learning models by quantifying the contribution of each feature to the predictions. The SHAP bar plots show the mean absolute impact of each feature on model predictions. Each bar's length indicates the overall importance, and the colors within each bar reflect the specific contributions to different maturity clusters. Figure 1.3[2] shows the overall feature importance (mean |SHAP| value) in the classification models. The length of each bar in Figure 1.3 represents how much that feature, on average, affects the prediction.

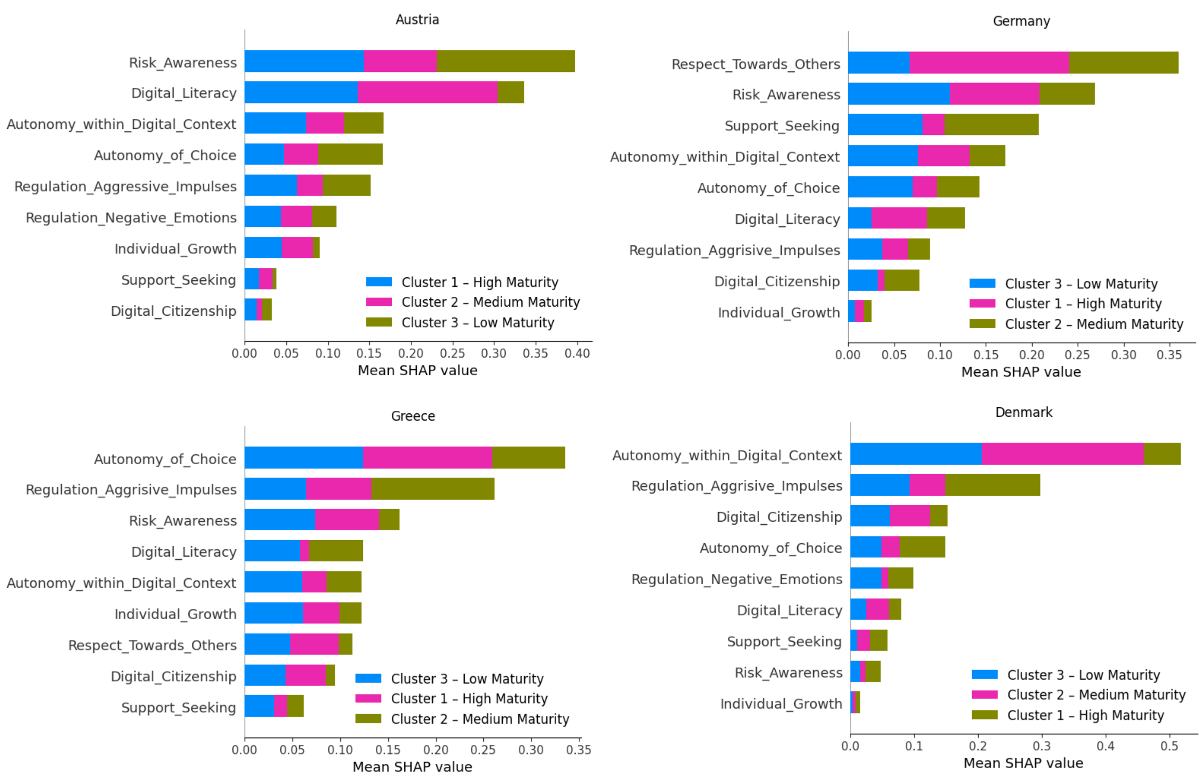

Figure 1.3: SHAP Values Obtained from Austria, Germany, Greece and Denmark

In Austria, the key predictors influencing digital maturity are primarily Risk Awareness, Digital Literacy, and Autonomy within Digital Context. Risk Awareness is significant across

[2]Note: Color-to-class mapping differs by country; consult each panel's legend. Bars show mean |SHAP| for one-vs-rest classifiers of high/medium/low maturity.

all maturity clusters, with particularly strong associations with both high and low maturity groups. This indicates that adolescents who clearly recognize online risks and proactively tend to demonstrate higher digital maturity, while those unaware or indifferent to these risks remain at lower maturity levels. Digital Literacy similarly affects maturity, significantly impacting the classification of individuals as highly mature, highlighting that technical proficiency and digital knowledge play crucial roles in shaping positive digital behaviors. Autonomy of Choice and Autonomy within Digital Contexts also exhibit considerable influence, suggesting that individual decision-making and responsible online navigation critically affect maturity.

In Germany, the top predictor is Respect Towards Others, highlighting the crucial role of civility and empathy in online interactions across all maturity levels. It is notably prominent in distinguishing high maturity, implying that youth with moderate digital maturity levels significantly differ based on their interpersonal digital conduct. Risk Awareness and Support Seeking follow closely, indicating that awareness of online dangers and proactive help-seeking behaviors substantially differentiate adolescents' maturity levels. Interestingly, Autonomy within Digital Contexts, Digital Literacy, and Regulation of Aggressive Impulses are influential yet less prominent compared to Austria, suggesting distinct cultural emphases in digital maturity competencies.

The Greek dataset highlights Autonomy of Choice as the most impactful predictor, significantly associated with all maturity clusters, particularly distinguishing high and low maturity individuals. This suggests that Greek adolescents who actively and consciously decide their digital engagements demonstrate elevated digital maturity. Regulation of Aggressive Impulses is another critical dimension, prominently affecting adolescents at all maturity levels, emphasizing the importance of behavioral self-control in online interactions. Additionally, Risk Awareness, Individual Growth, and Autonomy within Digital Context also significantly impact maturity, albeit with varied prominence, pointing towards a more balanced emphasis on personal decision-making and emotional-behavioral regulation.

For Denmark, Autonomy within Digital Context dominates the SHAP plot, impacting adolescents across all maturity levels, especially distinguishing medium and low maturity. This suggests Danish adolescents' capability to independently navigate digital platforms responsibly significantly correlates with better digital maturity. Regulation of Aggressive Impulses is highly influential, especially for high maturity group, highlighting the role of emotional control in online environments. Moreover, Digital Citizenship and Autonomy of Choice notably differentiate adolescents' maturity levels, emphasizing a broader civic engagement and individual decision-making approach to digital behaviors in Denmark.

Moreover, across all four countries, certain patterns and unique distinctions emerge. Interestingly, Risk Awareness consistently appears as a vital predictor in Austria, Germany and Greece, emphasizing its broad salience in shaping digital maturity irrespective of cultural differences. Autonomy (of Choice and within Digital Contexts) repeatedly emerges as significant, highlighting adolescents' independent decision-making capacity as critical for achieving higher maturity levels. Distinct country-specific emphasis on interpersonal dimensions like Respect Towards Others in Germany and Regulation of Aggressive Impulses in Denmark and Greece indicates cultural nuances do shape digital maturity competencies among children. For example, Digital Literacy, although crucial in Austria, appears relatively less prominent in other countries, suggesting variable perceptions and educational emphases on technical digital skills.

1.4.2 Why do These Features Dominate

The prominence of *risk awareness*, *autonomy* (of choice/within context), and *regulation of aggressive impulses* across countries (Figure 1.3) can be read through both a psychosocial-development lens and a socio-technical lens. Developmentally, these dimensions align with the triad of *temperance*, *perspective*, and *responsibility* that underpins maturity of judgment (Greenberger and Sørensen, 1974). In particular, self-regulatory control of impulses and emotions supplies the *brake*, while autonomy provides the *steering* that turns digital opportunities into growth rather than harm. In our context, regulation of digital impulses instantiates temperance, risk awareness reflects perspective on consequences, and autonomous technology use expresses responsibility. These capacities typically strengthen through adolescence, which helps explain why higher scores on impulse control, risk consciousness, and autonomy track with more mature digital engagement (Greenberger and Sørensen, 1974). Empirically, the ability to self-regulate online behavior (e.g., restraining aggressive impulses or managing negative affect during digital interactions) reliably emerges as a cornerstone of digital maturity. Adolescents with stronger self-control are less likely to engage in risky online behaviors, whereas those with weaker control report more negative outcomes Laaber et al., 2023). Autonomy, in turn, is linked to more instrumental, goal-directed uses of technology and to proactive help-seeking when problems arise, reinforcing other facets of maturity. In short, impulse regulation and autonomy work in tandem as self-regulation keeps behavior in check, while autonomy allows adolescents to translate digital opportunities into constructive action. This synergy is consistent with recent structural work showing that regulation of aggressive impulses occupies a central, connective position among digital competencies, with potential spillovers to other domains (Hofmans et al., 2024).

A socio-technical lens explains why these same features dominate across contexts yet vary in relative salience. Digital maturity is not a single trait but a *capability configuration* that develops in response to the *affordances and constraints* of technologies and local routines (Leonardi, 2011). Affordance theory emphasizes that outcomes depend less on the artifact itself than on how users perceive and enact what it enables or forbids; the same feature (e.g., anonymous messaging) can be construed as a hazard to be navigated cautiously by a risk-aware, self-regulated adolescent or as an invitation to impulsive behavior by a less mature peer. This helps explain our cross-country patterns where *risk awareness* surfaces as universally salient in Austria, Germany, and Greece, whereas *autonomy within digital context* is most dominant in Denmark, plausibly reflecting differences in school practices and youth norms that place stronger daily demands on independent, self-guided use (Livingstone and Helsper, 2007; Helsper et al., 2013). Put differently, each context sends a ‘demand signal’ for certain competencies where exploration and independence are emphasized, autonomy becomes especially distinguishing, and where exposure to online risks is higher or more salient, vigilance and prosocial restraint (including respect toward others) become sharper differentiators. These patterns are theoretically expected rather than methodological artifacts, and they reinforce the view that adolescents cultivate the mix of psychosocial skills like self-restraint, judicious risk assessment, and independent agency that are needed to thrive within the specific affordances and constraints of their digital milieu (Livingstone and Helsper, 2007;

Helsper et al., 2013).

Moreover, the observed cross-country contrasts in SHAP rankings (e.g., respect toward others in Germany; autonomy within context in Denmark) may not represent contradictory findings instead, these differences can be interpreted as evidence that the same underlying set of digital capabilities is differentially mediated by national institutional and cultural contexts. The EU Kids Online (Smahel et al., 2020) documents that exposure to online opportunities and risks co-vary, i.e. countries with stronger skills and autonomy often also register more risk exposure, which requires stronger risk awareness to keep the harm low (Livingstone and Helsper, 2007; Livingstone et al., 2011, 2018). This helps explain, for example, why German high maturity students show only average risk awareness while excelling in pro-social conduct. Similarly, Danish emphasis on autonomy within digital context alongside comparatively weaker digital citizenship maps onto educational systems that prioritize independent navigation of technology over civic uses of platforms. Country-specific SHAP vectors should be read as local demand signals for capability-building rather than as contradictions. This turns cross-country heterogeneity into a targeting logic. For example prioritize risk awareness where safety schemes may be weak, translate autonomy into citizenship where civic behaviors might lag, and sequence supports accordingly in schools, platforms, and ministries.

1.4.3 Classification Performance and Benchmarking Across Countries

All models achieved high classification performance, indicating that children's digital maturity levels can be predicted with strong accuracy using behavioral features. Table 1.4 summarizes the test performance of the full-feature ensemble (RF+DNN on all features), the top-5 feature ensemble, and the DRL model for each country. Overall, the RF+DNN full model attained the baseline highest accuracy and F1-score in each country when using the features. However, the DRL agent (Top-3 + DNN) matched or slightly exceeded this performance in every case despite using only three features (plus the DNN's aid). The Top-5 feature ensemble performed slightly worse than the full model in Austria, Germany, and Greece, and slightly above it in Denmark, but not far behind, highlighting that much of the predictive signal is captured by the five key features.

- **Austria:** The full RF+DNN classifier achieved an accuracy of 92% on the Austrian test set, with a F1-score of 0.923. The Top-5 feature model yielded about 82% accuracy (F1≈0.822), indicating a small, but significant drop due to feature reduction. The DRL model obtained 95% accuracy with an F1-score of 0.955, outperforming the full-model ensemble. In practical terms, the DRL agent misclassified only around 5% of Austrian children, correctly labeling Low/Medium/High maturity with near 95% success. This is notable given the DRL used significantly less information than the full model. Statistical tests confirmed that the DRL's improvement in F1 over the Top-5 model is significant ($p < 0.01$), and its slight edge over the full model is marginally significant (p $\approx$ 0.08), suggesting a positive trend.

- **Germany:** Results in Germany followed a similar pattern. The RF+DNN full model reached 81% accuracy (F1≈0.81) on German data. The Top-5 model obtained 80%

accuracy (F1≈0.80). The DRL model attained 90% accuracy with F1≈0.902, again slightly higher than the full model. We note that Germany had a smaller sample than Austria, which can make the classification task slightly harder. A paired t-test across the repeated trials indicated the DRL's improvements over the full model in Germany were not statistically significant ($p > 0.05$). Both the full and DRL models significantly outperformed the Top-5 model in Germany ($p < 0.01$), confirming that using only five features incurred a modest but real decline in predictive power.

- **Greece:** In Greece, overall performance was slightly lower, which may be attributed to the smaller sample size. The full model achieved 81% accuracy (F1≈0.81) on the Greek test set. The Top-5 ensemble saw accuracy drop to 74% (F1≈0.74). The DRL model, by contrast, reached 94% accuracy (F1≈0.94), again higher than the full model. Notably, in Greece the DRL's advantage was more pronounced. This suggests the DRL agent utilized the DNN's knowledge effectively to compensate for the limited data, arguably by generalizing patterns appearing in the dataset. All differences here were statistically significant (the DRL vs. full model $p < 0.05$, full vs. top-5 $p < 0.01$).

- **Denmark:** The Danish test results mirrored Austria's high performance. The RF+DNN full model attained 80% accuracy (F1≈0.799). The Top-5 model scored 82% accuracy (F1 ≈ 0.825). The DRL agent achieved about 94% accuracy (F1 ≈ 0.94), once again, slightly exceeding the full model. We observed that Denmark's data yielded very clear separation between the three maturity groups, perhaps because Danish children's digital habits were more homogeneous, making classification easier. In fact, the high maturity group in Denmark was particularly distinct (characterized by consistently high scores on all dimensions), which all models captured well (precision and recall for High class > 95%). The difference between DRL and full model in Denmark was statistically significant, however the Top 5 model and the DNN model accuracies were not found to be statistically significant.

Table 1.4: Classification performance by model and country.

Country	**RF+DNN (all)**	**Ensemble (Top-5)**	**Hybrid DRL (Top-3 + teacher)**
Austria	Acc 0.92, F1 0.92	Acc 0.82, F1 0.82	**Acc 0.95, F1 0.95**
Germany	Acc 0.81, F1 0.81	Acc 0.80, F1 0.80	**Acc 0.90, F1 0.90**
Greece	Acc 0.81, F1 0.81	Acc 0.72, F1 0.71	**Acc 0.94, F1 0.94**
Denmark	Acc 0.80, F1 0.79	Acc 0.82, F1 0.82	**Acc 0.94, F1 0.93**

Full-feature DNN probabilities are used only during training; deployment requires only the Top-3 raw features. **Bold** marks the best F1 within each country.

The DRL model consistently outperformed the full feature and Top-5 feature ensemble in accuracy and F1 in all four countries thereby confirming that the reinforcement learning approach adds significant value when only limited features are available, effectively recovering performance that would otherwise be lost. When comparing DRL vs. the full RF+DNN model, we found that DRL had a slightly higher mean F1 in all countries. The full model vs. Top-5 model comparison showed that using all features gives a small but significant boost in each country. This indicates that while the five most important features carry most of the information, the remaining features (e.g. the other DIMI dimensions) collectively still contribute a non-trivial amount to classification which is enough to slightly improve accuracy when included.

Performance of the DRL Model: The DRL agent's success with reduced input dimensions has important implications. It achieved comparable and even superior performance to the full-feature model, demonstrating that a small set of behavioral features is sufficient to assess digital maturity with high accuracy when paired with an intelligent decision policy. The top-3 features used by the DRL were consistently the core drivers of model predictions. For instance, the agent heavily utilized the Emotional Self-Regulation score in its decisions, i.e, children with low Regulation of Negative Emotions were rarely classified as high maturity by the distilled policy, consistent with theory.[3] In contrast, high self-regulation virtually guaranteed a classification of at least Medium or High. This aligns with the theoretical understanding that managing one's negative emotions online is a key component of digital maturity (Arenas and Yazdi, 2022).

Cross-Country Comparison: Our findings were largely consistent across Austria, Germany, Greece, and Denmark, reinforcing the generalizability of the digital maturity classification approach. Each country's data yielded three clusters of similar nature (High, Medium, Low maturity) with analogous characteristics, suggesting that the underlying construct of digital maturity manifests similarly despite cultural and contextual differences. Performance metrics were uniformly high (all models in all countries had accuracy well above 80% and F1 above 0.80), which speaks to the robustness of the model and the relevance of the chosen features across populations. We observed minor performance variations; Austria, Greece and Denmark had the highest accuracies, while Germany had the lowest. Despite the differences, the rank-order of model performance (DRL > Full > Top-5) held true in all four contexts.

The analysis also revealed differences in the top predictive features across the four countries, highlighting the culturally specific nature of digital maturity predictors. For Austria, the primary predictors included Risk Awareness, Digital Literacy, and Autonomy within Digital Context, emphasizing the importance of safety-conscious behavior, digital skills proficiency, and voluntary engagement and dis-engagement with technology. Germany's key indicators were different, highlighting Respect Towards Others, Risk Awareness, and Support Seeking, suggesting an emphasis on social interaction, safety awareness, and proactive help-seeking behavior online and offline. In Greece, Autonomy of Choice, Regulation of Aggressive Impulses, and Risk Awareness emerged as critical, highlighting voluntary digital engagement, behavioral self-regulation, and online risk consciousness. Denmark prioritized Autonomy within Digital Context, Regulation of Aggressive Impulses, and Digital Citizenship, indicating an emphasis on responsible technology use and ethical digital behavior. Importantly, Risk Awareness appeared consistently among the top features for Austria, Germany, and Greece, emphasizing its universal significance in predicting digital maturity. This consistency implies that despite cultural and educational contexts, awareness and understanding of online risks fundamentally strengthen digital maturity behavior, aligning with previous literature which identifies risk consciousness as a crucial determinant of safe and responsible digital engagement (Livingstone and Helsper, 2007; Vandoninck et al., 2013).

[3] The teacher probabilities were used only during training to guide learning and are not required at deployment.

Robustness Checks

The robustness checks confirm the reliability and stability of our hybrid DRL model across datasets from Austria, Germany, Greece, and Denmark (Figure A2 . Sensitivity analysis revealed consistent and stable performance across multiple hyperparameter configurations. For all the country datasets, accuracy ranged from 67.7% to 92.8%, with optimal performance observed at moderate learning rates and larger batch sizes. This sensitivity analysis suggests that our hybrid DRL model maintains high performance within reasonable parameter ranges, highlighting model robustness and reliability. Evaluations using multiple random seeds (Figure A3) confirmed consistent predictive performance, highlighting the robustness of our hybrid DRL approach. The model achieved a mean accuracy of 92.2% (0.42%) and mean F1-score of 92.3% (0.41%) in Austria, with similarly stable performances in Germany (accuracy 81.9% 1.01%, F1 82.1% 1.06%), Greece (accuracy 81.3% 0.00%, F1 81.6% 0.06%), and Denmark (accuracy 77.5% 0.00%, F1 77.0% 0.00%). The minimal variance across repeated runs confirms the reproducibility of results and reliability of the modeling framework regardless of random initialization. The ablation study demonstrated (Figure A4) the crucial contribution of integrating DNN-generated probabilities into the DRL model. Excluding DNN probabilities resulted in substantial accuracy declines, highlighting their significant informational value. Specifically, Austria experienced a sharp drop from 92.2% to 70.1% accuracy, Germany declined from 83.2% to 62.8% accuracy, Greece from 81.3% to 68.0% accuracy, and Denmark from 82.5% to 70.0% accuracy . Thus, the inclusion of full-dimensional DNN probabilities significantly enhanced model performance, providing improvements of approximately 22.1% in Austria, 21.3% in Germany, 12.8% in Greece, and 17.8% in Denmark (F1-scores). These findings validate the hybrid transfer learning approach.

1.4.4 Contributions and Implications

This study advances the academic discourse and practice on adolescent digital maturity by making contributions that bridge theoretical constructs with operationally viable assessment tools. First, this study contributes to the digital maturity literature by translating a multi-dimensional, psychology informed construct into a compact, interpretable measurement and decision system that can be embedded in school and platform operations. Specifically, we align the Digital Maturity Inventory (DIMI) and its antecedents in psychosocial maturity (Greenberger and Sørensen, 1974) with model-based attributions that pinpoint which competencies (e.g., risk awareness, autonomy of choice/within context, regulation of aggressive impulses) most strongly drive maturity classification. SHAP analyses elevate these dimensions across countries, thereby providing a theoretically grounded basis for targeted action rather than undifferentiated 'screen-time' prescriptions (Livingstone et al., 2011; Laaber et al., 2023). This integration complements recent empirical work on the developmental structure of digital maturity (Hofmans et al., 2024) by specifying which facets most shift predicted maturity, for whom, and in what direction. This level of granularity is necessary for practice in schools and youth-facing services. In doing so, we respond to long-standing calls in Operations Management and IS research to move from abstract constructs to scalable decision aids (Benbya et al., 2020).

In terms of operational meaning of 'Value of Information', our DRL algorithm, trained

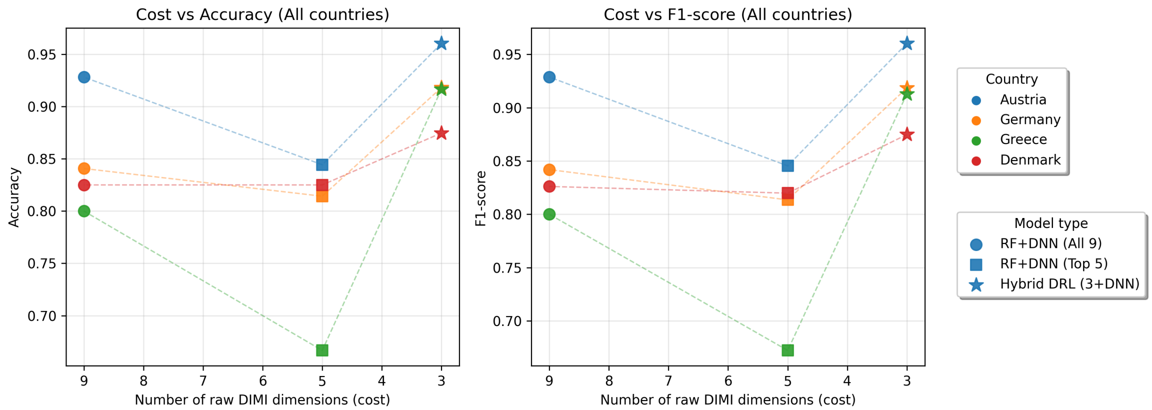

Figure 1.4: Pareto-Improving Operating Point

with distilled teacher signals and using only three raw features at deployment, matches or exceeds the full-feature ensemble. This is an operations result. The agent finds a *Pareto-Improving Operating Point* (Figure 1.4) on the accuracy–cost frontier by acquiring only high-value features and transferring knowledge from the full model (Janisch et al., 2020; Mnih et al., 2015). For operations management literature, this shows interpretability and cost-optimality are complements, i.e., by aligning SHAP-identified levers with a cost-aware policy, decision makers can reduce survey load while preserving transparency for governance (Doshi-Velez and Kim, 2017; Rudin, 2019). Two research implications follow:

- **Diminishing returns to additional features:** The Top-5 model's slight drop versus the full model, combined with DRL's recovery of accuracy. This shows that the extra features offered very little additional benefit, a pattern known as diminishing marginal returns. This supports the idea that intelligent, sequential decision-making policies are the best approach for this type of problem (Janisch et al., 2020).
- **Knowledge distillation as an Operations Management instrument**. The ablation study shows large performance losses when distillation inputs are removed, implying that distillation converts expensive multi-item measurement into a reusable informational asset that lowers future operating costs without degrading decision quality (Hinton et al., 2015).

1.4.5 Practical Implications

Schools can apply these findings through a lightweight triage-and-support system that integrates cost-efficient screening with targeted interventions. Using the study's three most predictive features for each local context, schools can implement a short 'micro-screen' that takes minimal class time yet accurately identifies students at low, medium, or high digital maturity. Those flagged for support can receive brief, competency-specific interventions, for example, a five-minute risk-recognition activity, impulse-control exercises for online communication, or visible help-seeking channels. At the school level, aggregated SHAP explanations can populate a simple dashboard that helps principals allocate limited counseling minutes, prioritize teacher professional development toward the most elastic competencies, and track improvement over time as part of routine cycles. In short, interpretability plus cost-aware screening converts measurement into day-to-day operational decisions about who needs what support, when, and with which instructional lever.

Moreover, platforms can translate the interpretable feature insights into 'safety-by-design' measures. Low autonomy or high impulsivity profiles could trigger 'are you sure?' friction prompts before content is posted, while low support-seeking scores could surface quick-access help buttons. Adaptive information collection like asking only the most predictive few questions during onboarding, can personalize safety tools without overburdening the user, especially when combined with offline-trained knowledge-distilled models to preserve privacy. These measures not only improve user safety but also help platforms meet audit and transparency requirements for youth-facing AI systems.

Furthermore, policymakers can embed the interpretable digital maturity index into national education dashboards, directing resources toward the competencies with the highest marginal returns. Governments can run annual full-feature surveys to refresh the Hybrid DRL 'teacher model' that informs the three-item screen. This targeted approach scales coverage, focuses funds where they matter most, and ensures interventions are responsive to changing patterns in adolescent digital behavior.

1.5 Conclusion

This study demonstrates that adolescents' digital maturity can be predicted with high accuracy from a small, carefully selected set of psychosocial and behavioral indicators. Across four European countries, we identify three consistent maturity profiles and show that a heterogeneous ensemble (RF + DNN) delivers strong predictive performance. More importantly, a cost-aware hybrid DRL policy matches or exceeds this performance using only three features, reducing measurement burden by two-thirds. SHAP explanations reveal a common set of actionable levers (risk awareness, autonomy (of choice/within context), and impulse regulation) that can guide targeted interventions in schools, platforms, and policy programs. By combining interpretability with operational efficiency, we offer a scalable, theory-aligned decision system that links the construct of digital maturity directly to implementable actions.

This work is not without limitations, though in several cases we have taken steps to mitigate them. First, our maturity labels originate from unsupervised clustering rather than externally validated ground truth. While this means they remain proxies, we addressed this by using a principled clustering method (GMM with BIC selection) and by grounding the resulting clusters in established developmental theory. The next step is field validation against independent behavioral and well-being outcomes, which would strengthen the maturity construct as a policy-relevant performance measure. Second, our data capture a single point in time. Digital maturity is inherently developmental, shaped by experience, institutional context, and technological change. We partly offset this by designing a low-burden, three-item screening approach that can be deployed at high frequency, enabling longitudinal monitoring in practice. Future research should take advantage of this capability to run adaptive field experiments that can test not just whether interventions work, but how quickly and for whom they generate change. This paper offers both a method and a blueprint, a way to move from a rich theoretical construct to a practical, interpretable, and cost-optimized decision system that is ready for operational use. The results suggest that a small number of well-chosen dimensions can provide reliable guidance for resource allocation in education and platform governance, without sacrificing transparency or efficiency. The challenge ahead is to embed such systems into real-world feedback loops so that the prediction of digital maturity is not an end in itself, but the starting point for measurable improvement in how young people navigate and thrive in their digital worlds.

Decoding Digital Maturity: An Explainable AI Approach to Adolescent Development

Abstract

Children's digital maturity, defined as the skills, self-regulation, and awareness that enable young people to use technology in ways that support rather than undermine their development, is known to vary across adolescence. Yet research has largely treated these capabilities as stable individual differences, overlooking how their relative importance shifts as young people mature and whether these shifts differ by gender. We propose that digital maturity development is phase-structured, that is, different dimensions become rate-limiting at different developmental stages, and the binding constraints may differ systematically by gender. We test this proposition using a two-layer analytic approach combining flexible machine learning for pattern discovery with formal statistical testing for confirmation. Layer one employs Random Forest regression with explainable AI (SHAP) to identify salient dimensions and characterize functional forms. Layer two applies confirmatory regression models with multiple-testing corrections to evaluate statistical reliability. Analyzing responses from 3,018 adolescents aged 11–18 across Austria and Germany, three findings emerge. Digital Literacy operates as a universal foundation in early adolescence, confirming across subgroups, but its salience fades as populations achieve baseline competence. Emotional regulation emerges as uniquely important for older female adolescents—a pattern replicating across both countries but absent for males—consistent with an affordance-vulnerability mismatch where platform features place differential regulatory demands on users with gendered socialization patterns. Within-age heterogeneity is substantial at the individual level, with developmental positions spanning two to three years among same-age peers, though average differences between maturity groups are more modest. These findings contribute to IS research as they challenge static conceptualizations of digital capabilities by demonstrating phase-structured development. It also extends affordance theory by showing how platform features interact with developmental stage and gendered vulnerabilities and this research documents heterogeneity that complicates assumptions underlying age-based digital governance frameworks. Practically, findings suggest that digital education should be calibrated to developmental phase, platforms should support emotional regulation for vulnerable users, and policymakers should recognize that age-based thresholds systematically often misclassify adolescents at developmental extremes.

Key Words: Digital Maturity, Explainable AI, Adolescent Development, Digital Inequality, Gender Differences

2.1 Introduction

The rapid proliferation of Information and Communication Technology (ICT) devices, including smartphones, tablets, and laptops, has fundamentally transformed how people access information, interact with peers, and learn. By the end of 2024, a staggering 5.8 billion mobile internet users were reported which is estimated to increase to 6.5 billion by 2030 (GSMA, 2025). This exponential growth in digital connectivity draws attention to the widespread influence of technology in shaping contemporary societies worldwide. While this digital revolution has had an impact on all social and demographic group, its impact on children and adolescents is particularly noteworthy (Vissenberg et al., 2022). Today's youth are growing up in a digital landscape where access to ICT devices is ubiquitous and digital skills are increasingly valued. As such, children are not merely passive consumers of technology but are actively engaged participants, often referred to as 'digital natives' (Livingstone et al., 2018). This paradigm shift in childhood experiences highlight the need to understand how children and adolescents navigate the digital world and develop the competencies necessary to thrive in an increasingly digitized society.

Information Systems (IS) research has made substantial progress in understanding digital capabilities—the skills and competencies that enable individuals to use technology effectively. Work on digital inequality demonstrates that disparities increasingly reflect differences in skills and usage patterns rather than mere connectivity (Van Deursen and Van Dijk, 2011, 2014; Hargittai, 2001). Research on technology adoption and use has examined how individual characteristics shape outcomes from IT engagement (Burton-Jones and Straub Jr, 2006; Marcolin et al., 2000). Yet this body of work shares a common, largely unexamined assumption: that digital capabilities are relatively stable individual differences, accumulated incrementally through training and experience, possessed by agents whose underlying cognitive and emotional capacities remain constant. This assumption may hold for adult populations but becomes problematic when applied to adolescents. Developmental science establishes that adolescence is a period of profound reorganization—cognitive control, reward sensitivity, and affective regulatory systems mature on different timelines, fundamentally altering the capacities available for navigating complex environments (Silvers, 2022; Steinberg, 2005). If the agent is itself changing, then which capabilities matter for effective digital engagement may also change. The 'skill' that enables beneficial technology use at age 12 may differ categorically from the capability required at age 17. This creates a theoretical puzzle that IS research has not adequately addressed which is essentially how should one conceptualize digital capabilities when the underlying cognitive, emotional, and social capacities of users are themselves developing?

We argue that for adolescent populations, digital capabilities should be understood not as stable individual differences but as phase-structured developmental processes where different factors become limiting at different stages. This reconceptualization has three implications that motivate the present study. First, if development is phase-structured, then some capabilities may operate as universal foundations which are necessary preconditions that must be established before later capabilities become salient. Second, because developmental trajectories are shaped by social context, the binding constraints on digital maturity may differ systematically by gender, reflecting how platform affordances interact with socialized vulnerabilities to create differential capability demands. Third, if capabilities develop at different rates across individuals, then adolescents of identical chronological age may occupy meaningfully different developmental positions, challenging assumptions that underlie age-based digital governance frameworks.

To investigate these possibilities, we focus on *Digital Maturity*, defined as the skills and perspectives that enable individuals to use digital technologies in ways that support individual development and social integration (Laaber et al., 2023). Digital maturity is not a single skill, but a multidimensional capability set. Recent empirical work shows that digital maturity is meaningfully associated with outcomes that societies care about, including aspects of digital well-being and mobile device use behaviors (Laaber et al., 2024). It encompasses not only technical competencies such as digital literacy but also self-regulatory capacities such as emotional regulation and impulse control, as well as prosocial orientations such as respect toward others and digital citizenship. This multidimensional structure makes digital maturity well-suited for examining whether different dimensions become salient at different developmental stages and for different demographic groups. From an IS perspective, the dimensional structure of digital maturity has implications for intervention design, measurement validity, and theory-building about capability development in sociotechnical environments. If a small subset of dimensions is consistently central across subgroups, that suggests a stable core of digital capability. If other dimensions shift in importance by age or gender, that points to context-sensitive developmental pathways and to potential boundary conditions for theories that treat digital competence as uniform across youth populations. The study addresses two research questions:

RQ1: How does the relative importance of digital maturity dimensions vary across adolescent development, and do some dimensions serve as foundational preconditions for later development?

RQ2: How do gender and individual developmental heterogeneity shape digital maturity trajectories among adolescents?

To address these questions, we employ a hybrid analytic strategy that combines flexible machine learning for pattern discovery with formal statistical testing for confirmation. First, we use ensemble learning to learn associations between digital maturity dimensions and age, and then interpret those learned associations using additive feature attribution methods. Second, we complement these explanations with spline-based confirmatory models and joint significance tests, allowing us to formally evaluate whether the patterns highlighted by SHAP correspond to statistically distinguishable, potentially non-linear relationships. This design responds to long-standing concerns about conflating exploratory and confirmatory goals (Shmueli, 2010) and reflects growing recognition that prediction-oriented approaches can complement explanation by identifying patterns that traditional linear specifications may miss (Shmueli and Koppius, 2011). We analyze survey data from 3,018 adolescents aged 11–18 across Austria and Germany, enabling both within-country subgroup analysis and cross-national comparison. The contributions are threefold. Conceptually, the study advances digital maturity research by shifting attention from aggregate maturity scores to the relative salience and shape of the underlying dimensions across adolescence. Methodologically, it demonstrates a reproducible hybrid workflow that combines flexible prediction (random forests) with transparent explanation (SHAP) and confirmatory modelling of non-linear marginal effects. Empirically, it offers evidence across adolescent subgroups and settings to clarify which dimensions appear robust versus contingent, thereby informing both future theory development and practical targeting of educational or policy interventions.

The study makes three contributions to IS research. First, we advance theory on digital capabilities by introducing the concept of phase-structured digital capability development. We demonstrate empirically that Digital Literacy operates as a universal foundation in early adolescence but that its salience diminishes as populations achieve baseline competence, revealing later-stage binding constraints related to self-regulation. This challenges static conceptualizations of digital skills and extends frameworks on digital inequality by showing that the "binding constraint" on effective use shifts across development. Second, we extend affordance theory by demonstrating how platform features interact with users' developmental stage and gendered vulnerabilities to create differential capability demands. Specifically, we find that emotional regulation emerges as uniquely important for older female adolescents (a pattern that replicates across both national contexts but is absent for males). We interpret this through a sociomaterial lens as visual social media platforms afford continuous social comparison opportunities that may place distinctive regulatory demands on users socialized toward relational orientation and emotional expressivity. This finding introduces the concept of affordance-vulnerability mismatch and suggests that affordance analysis for adolescent populations must explicitly consider developmental stage and socialized vulnerabilities. Third, we document within-age heterogeneity in developmental status that is substantial at the individual level (predicted developmental positions spanning two to three years among same-age peers) yet modest on average (approximately four months between high and low-maturity groups). This finding challenges assumptions underlying age-based digital governance frameworks while avoiding overly strong claims. Age-based thresholds may provide reasonable approximations for typical adolescents but may systematically misclassify those at developmental extremes, both advanced adolescents who may be unnecessarily restricted and delayed adolescents who may be insufficiently protected. We also examine whether digital maturity relates to weekly usage intensity as a robustness check, finding that the maturity captured by the model is not simply a proxy for time spent online or intensity or use.

2.2 Literature Review

2.2.1 Digital Maturity as a Multi-Dimensional Concept

Previous research on young people's ICT use has primarily centered on screen time and basic digital literacy skills overlooking the broader social and cultural contexts that shape their digital engagement (Livingstone et al., 2018; Reid Chassiakos et al., 2016; Hill et al., 2016). While excessive screen time can have negative effects (Van Deursen and Van Dijk, 2014) like depression (Boers et al., 2019), digital literacy remains a crucial enabler for safe and effective technology use (Livingstone et al., 2018). However, there is a growing recognition of the need for a more comprehensive framework. Recent studies emphasize the importance of understanding young people's ICT use within the social and cultural environments where they interact with technology (Livingstone et al., 2023). In response to this call, the concept of Digital Maturity has emerged as a more holistic approach to understanding young people's digital skills and competencies (Laaber et al., 2023). Digital maturity extends beyond technical skills and screen time management focusing on young people's ability to assess and regulate their behavior with technology in different contexts. By developing digital maturity, young people gain the tools to navigate the complexities of the digital world and make informed choices about their online engagement (Laaber et al., 2024).

Digital maturity is a modern concept offering a view on how young people use technology in today's world. According to Laaber et al. (2023), digital maturity is operationalized through ten dimensions organized around three core competency areas. The first area concerns the ability to use technology independently and make autonomous choices, captured through Autonomous Choice to Use Mobile Devices and Autonomy Within Digital Contexts. The second area involves overcoming challenges and solving problems in digital environments, encompassing Digital Literacy, Individual Growth in Digital Contexts, Digital Risk Awareness, and Support-Seeking Regarding Digital Problems. The third area addresses interacting appropriately with others and contributing positively online, including Regulation of Negative Emotions in Digital Contexts, Regulation of Impulses in Digital Contexts, Respect Towards Others in Digital Contexts, and Digital Citizenship. This multidimensional structure matters theoretically because a purely technical framing cannot explain why two adolescents with similar access and similar basic skills may nevertheless differ sharply in safety outcomes, conflict escalation, susceptibility to manipulation, or the ability to recover from negative digital experiences. The capabilities that protect adolescents in digital environments often resemble self-regulation and socio-emotional capacities expressed under digital affordances, not merely operational ability. Recent empirical work demonstrates that digital maturity is meaningfully associated with outcomes societies care about, including aspects of digital wellbeing and mobile device use behaviors (Laaber et al., 2023).

IS research on digital inequality provides a complementary perspective.Van Dijk's (2005) influential framework proposes sequential stages of digital inequality as motivational access, material access, skills access, and usage access. Hargittai's (2001) foundational work on the second-level digital divide demonstrated that access alone is insufficient—differences in skills and usage patterns create new forms of inequality. Van Deursen and Helsper (2015) extended this to a third-level digital divide focusing on outcomes from technology use. This progression from access to skills to outcomes creates a conceptual bridge to digital maturity as maturity can be understood as the set of competencies that allow individuals to convert access into beneficial outcomes while managing risks. However, these frameworks share an important limitation when applied to adolescent populations. They implicitly assume relatively stable agents progressing through stages. The frameworks describe what capabilities matter, but not how those capabilities themselves develop or whether different capabilities matter at different developmental stages. For adolescents, whose underlying capacities are themselves rapidly developing, this assumption may be untenable.

2.2.2 Adolescent Development and the Changing Demands of Digital Engagement

Adolescence is not a smooth linear ramp toward adulthood but rather a period of discontinuous development characterized by different systems maturing on different timelines.

Developmental science establishes that cognitive control, reward sensitivity, and affective regulatory systems follow distinct developmental trajectories, creating periods where regulatory demands may exceed regulatory capacity for many youth (Steinberg, 2005). Emotion regulation in particular undergoes substantial development across adolescence, with capacities continuing to mature into early adulthood (Silvers, 2022; Zimmermann and Iwanski, 2014). This asynchronous development has implications for how adolescents navigate complex environments, including digital ones. Empirical research documents that digital activities change substantially across adolescence. Younger children primarily engage with digital devices for interactive play and exposure to age-appropriate educational materials (Marsh et al., 2020). As children transition into adolescence, their technology use becomes more diversified and integrated into various aspects of their lives. Teenagers exhibit a broader range of digital activities, including social media engagement, online communication, gaming, and content creation (Vogels et al., 2022). These changing patterns of engagement imply changing demands on users' capabilities. Gender also plays a significant role in shaping digital engagement patterns. While many gender differences are small on average, they can become meaningful under specific social and situational contexts (Hyde, 2005; Liang et al., 2023). Studies consistently show that girls tend to gravitate toward social networking and communication platforms, using technology primarily for maintaining social connections and engaging in online conversations (Ma, 2022; Twenge and Campbell, 2018). Boys demonstrate stronger inclination toward gaming and multimedia content (Twenge and Campbell, 2018). These different usage patterns suggest that boys and girls may encounter different challenges and develop different capabilities as they navigate digital environments.

Taken together, developmental science and research on youth digital engagement suggest that the relationship between digital maturity dimensions and developmental stage is unlikely to be uniform. If different systems develop on different timelines, and if different activities place different demands on those systems, then which maturity dimensions are most relevant may shift across adolescence. Early adolescence, when digital literacy skills are still developing and usage patterns are less complex, may be dominated by foundational competencies. Later adolescence, when social media engagement intensifies and self-regulatory demands increase, may see emotional and impulse regulation become more important. This possibility that digital capability development is phase-structured rather than uniform motivates our first research question. We propose that digital maturity development follows a phase-structured pattern characterized by two key features: foundational capabilities that operate as universal preconditions, and phase-specific constraints that shift across development. The logic of foundational capabilities draws on skill acquisition research demonstrating that complex competencies typically build on simpler precursors. Just as reading comprehension requires basic decoding skills, sophisticated digital engagement may require foundational technical competencies. Without basic literacy, adolescents cannot effectively access digital environments where higher-order capabilities become relevant. However, foundational capabilities are characterized by ceiling effects. Once populations achieve baseline competence, variation in that capability ceases to discriminate developmental status, not because the capability no longer matters, but because it is no longer the binding constraint. This pattern has been documented in digital literacy research showing that gender differences in ICT literacy emerge in early adolescence but diminish by later adolescence as populations converge toward competence (Gnambs, 2021; Fraillon et al., 2020). As foundational capabilities are established, later stage binding constraints may also emerge. Developmental science suggests that these constraints should relate to self-regulatory capacities, precisely those capabilities that continue developing through later adolescence. Prosocial behavioral factors are all capacities that develop on extended timelines(Silvers, 2022). (Silvers, 2022). If digital environments place increasing regulatory demands as social media engagement intensifies, these self-regulatory dimensions should become more important in differentiating developmental status among older adolescents.

2.2.3 Gender, Platform Affordances, and Differential Capability Demands

The phase-structured account predicts that later adolescence will see self-regulatory dimensions become more important, but it does not specify whether these binding constraints will be uniform across demographic groups. We argue that they will not—specifically, that platform affordances interact with gendered socialization patterns to create differential capability demands. Sociomateriality theory emphasizes the mutual constitution of technology and social structures(Leonardi, 2012; Orlikowski, 2007). Digital platforms are not neutral tools providing equivalent affordances to all users. Rather, platforms afford different action possibilities and pose different challenges depending on users' socialized orientations, developed competencies, and social contexts shaped by gender and other structural positions (Faraj and Azad, 2012; Treem and Leonardi, 2013). Theories of technology affordances in IS have examined how platform features enable and constrain action possibilities (Majchrzak and Markus, 2012; Markus and Silver, 2008), but this work has largely focused on organizational contexts with adult users possessing mature cognitive and emotional capacities.

For adolescent populations, where regulatory capacities are still developing, affordance analysis must consider an additional dimension, the match or mismatch between platform demands and user capabilities. Visual social media platforms where adolescent females disproportionately engage (Anderson et al., 2018) afford continuous opportunities for upward social comparison on appearance, lifestyle, and popularity. Research robustly links such comparison opportunities to negative affect and psychological distress, particularly for females (Fardouly et al., 2015). Platform features quantify peer feedback through visible metrics and make relational dynamics hypervisible (Burrow and Rainone, 2017). These platform affordances interact with gendered socialization patterns. Girls are socialized toward communion, relational orientation, and emotional expressivity (Chaplin and Aldao, 2013; Rose and Rudolph, 2006). When users socialized toward relational sensitivity encounter platforms architected to amplify social comparison and quantify peer feedback, the result may be what we term an 'affordance-vulnerability mismatch' as platform features systematically trigger emotional responses requiring sophisticated regulation to manage adaptively. This creates differential capability demands, specifically, demands on emotional regulation capacities that may be particularly salient for adolescents navigating comparison-driven digital environments.

The concept of affordance-vulnerability mismatch differs from traditional person-technology fit frameworks (Goodhue and Thompson, 1995) in two important respects. First, it emphasizes that the relevant "person" characteristics are not stable individual differences but developmentally dynamic capacities. Second, it foregrounds the role of socialized vulnerabilities (not merely skills or preferences) in shaping how affordances are experienced. Where task-technology fit asks whether technology features support task performance, affordance-vulnerability mismatch asks whether technology features overwhelm regulatory capacities that are still developing. This reframes the design question from “does this feature help users accomplish goals?” to “does this feature place demands that exceed users’ developmental capacities?”

2.2.4 Within-Age Heterogeneity and the Limits of Age-Based Frameworks

IS research and digital policy commonly treat chronological age as a proxy for developmental status. Research on technology adoption has extensively examined age as a moderator of adoption and use (Morris and Venkatesh, 2000; Venkatesh et al., 2003). Regulatory frameworks use age-based thresholds like data collection prohibitions under age 13 (COPPA), platform age minimums, age-stratified content restrictions. These approaches assume that chronological age adequately captures relevant capabilities, needs, and vulnerabilities. If digital maturity is multidimensional and developmental trajectories vary across individuals, this assumption may becomes questionable. Person-technology fit frameworks suggest that effective use depends on alignment between user characteristics and technology demands (Goodhue and Thompson, 1995). For adults, relevant characteristics may be relatively stable individual differences. For adolescents, relevant characteristics are developing capabilities that may vary substantially among same-age peers—some advancing quickly, others more slowly. Two 14-year-olds may occupy meaningfully different developmental positions across multiple maturity dimensions. This creates an empirical question with significant policy implications as to what extent do adolescents of identical chronological age differ in their digital maturity profiles? If within-age heterogeneity is negligible, then chronological age provides a reasonable approximation of developmental status and age-based frameworks are defensible. If within-age heterogeneity is substantial, then age-based frameworks may systematically misclassify adolescents at developmental extremes. We thus approach within-age heterogeneity as an empirical question requiring both documentation (does heterogeneity exist?) and quantification (how large is it, both on average and for individuals at distributional extremes?). The findings will inform whether existing age-based frameworks require modest refinement or more fundamental reconsideration.

The theoretical framework developed above generates expectations at three levels. First, at the level of developmental phase, we expect some dimensions to operate as a foundational capability dominating early adolescence. Second, we expect prosocial factors to emerge as particularly important for older adolescents, reflecting the affordance-vulnerability mismatch created by visual social media platforms and gendered socialization patterns. Third, at the level of within-age heterogeneity, we expect variation among same-age peers. Testing these expectations requires a methodological approach that can identify which dimensions are most salient at different developmental stages, characterize the functional form of those relationships, assess whether identified patterns are statistically robust, and quantify within-age heterogeneity. Standard linear models are transparent but may misrepresent relationships involving thresholds, diminishing returns, or non-monotonic effects—precisely the patterns the theoretical framework anticipates. Algorithmic models can capture such structure but introduce interpretability risks that make theoretical conclusions difficult to defend (Rudin, 2019). We therefore adopt a hybrid analytic strategy that combines flexible machine learning for pattern discovery with formal statistical testing for confirmation. This design separates discovering structure from testing that structure, reducing concerns about overfitting explanations to data while maintaining the flexibility needed to detect non-linear relationships. The following section describes this methodological approach in detail.

2.3 Methodology and Data

The theoretical framework developed in Section 2 generates expectations about phase-structured development, gender-differentiated pathways, and within-age heterogeneity. To accomplish this without overstating causality, we use a hybrid analytic strategy that combines (1) flexible machine learning for discovery and pattern detection and (2) a transparent, inference oriented regression framework for confirmatory testing and interpretable effect shapes. This design aligns with long standing concerns about conflating explanatory and predictive modeling goals. Predictive models can be valuable for discovering complex nonlinear patterns, but explanatory claims require additional structure and inferential discipline (Shmueli, 2010; Breiman, 2001). The IS literature has similarly argued that prediction oriented approaches can complement explanation by helping identify patterns and boundary conditions that traditional linear specifications may miss (Shmueli and Koppius, 2011).

Operationally, we implement two layers, each producing a distinct type of evidence that feeds directly into the next stage and into the paper's results narrative. Layer 1 identifies which dimensions appear most strongly associated with age and provides an interpretable ranking and descriptive effect visuals. Layer 2 then subjects those candidate dimensions to formal statistical testing and yields smooth estimated effect curves with uncertainty, enabling more disciplined statements about which relationships are robust and whether they are linear, threshold like, or otherwise nonlinear.

- Layer 1 (Discovery): Random Forest models estimated using cross validation, where each observation's predicted age is generated by a model that was trained without that observation (out of fold prediction). We compute SHAP values on these held out predictions to identify which digital maturity dimensions are most strongly associated with age within each subgroup and to visualize heterogeneity in those relationships. The output of this layer is a ranked set of dimensions per subgroup, along with descriptive explanatory plots indicating the direction and potential nonlinearity of associations.

- Layer 2 (Confirmation): For each subgroup, we take the most important dimensions identified in Layer 1 and estimate semi parametric spline regressions that explicitly model potentially nonlinear relationships. We then conduct joint Wald tests on each spline block, using robust standard errors and false discovery rate adjustment, to evaluate whether the association between each candidate dimension and age is statistically distinguishable from zero in a way that is less sensitive to functional form assumptions. The output of this layer is a compact confirmatory panel of adjusted p values together with marginal effect curves and confidence bands, allowing us to characterize not only whether a dimension matters, but how it matters across the predictor range.

This sequencing is a disciplined version of explore then confirm where the discovery stage is explicitly separated (via out of fold estimation) from the confirmatory stage to reduce mechanical overfitting of explanations to the same data used to fit the predictive model. The logic is similar to sample splitting and cross fitting ideas used to improve validity when flexible models are used in early stages of analysis (Chernozhukov et al., 2018).

2.3.1 Data, Measures and Subgroups

For our analysis, we utilize data from the Digital Maturity Inventory (DIMI), a validated survey capturing diverse psychosocial and behavioral dimensions of adolescents' ICT usage (Laaber et al., 2023). This dataset was selected as it was originally collected to measure children's digital maturity and due to its comprehensive coverage of digital maturity constructs and international scope, allowing for robust cross-country comparisons. Our sample comprises adolescents aged 11–18 across two European countries: Austria (n=1721), Germany (n=1297) totaling 3,018 valid responses. Detailed demographics are presented in Table 2.1.

Table 2.1: Summary statistics by country

Statistic	**Austria**	**Germany**
N	1721	1297
Age		
Mean	14.67	14.82
SD	1.85	1.71
Gender		
Male	884 (51.37%)	638 (49.19%)
Female	828 (48.11%)	656 (50.58%)
Other	8 (0.52%)	3 (0.23%)

The predictors are the 10 digital maturity dimensions measured on a 5 point Likert type scale. The key outcome used in the current pipeline is chronological age (years). Gender is binary coded (Male, Female), and we operationalize developmental stage using two age bands (11 to 14 and 15 to 18). The reason of grouping the children in the mentioned age groups are as follows:

1. Development Differences: Children undergo significant cognitive, emotional, and social development during adolescence. By dividing them into age groups (11-14 years and 15-18 years), we acknowledge these developmental disparities and recognize that digital maturity may manifest differently across these stages. For instance, younger adolescents may exhibit different digital behaviors and competencies compared to older adolescents due to varying levels of cognitive maturation and exposure to technology.

2. Relevance to Educational Context: Age-based segmentation aligns with the educational structure prevalent in many countries, where schooling is typically categorized

into stages such as middle school (11-14 years) and high school (15-18 years). Understanding digital maturity within these educational contexts allows for insights into how technology integration in academic settings may influence digital competencies and behaviors.

3. Targeted Interventions: Segmenting children into age groups facilitates the identification of age-specific challenges and opportunities concerning digital maturity. This segmentation enables the tailoring of interventions and educational programs to address the unique needs and developmental milestones of children within each age bracket. For instance, interventions aimed at promoting digital literacy may adopt age-appropriate strategies and content to resonate with the cognitive abilities and interests of children in each age group.

To improve comparability across features and across plots, we standardize each digital maturity dimension using z scores (mean 0, SD 1) prior to modeling. Standardization does not change rank orderings of model based importance measures in any deterministic way, but it improves interpretability of marginal effect plots by placing predictors on a common scale. We also remove observations with missing values on any of the 10 predictors, age, or gender. We then create four subgroups: male 11 to 14, male 15 to 18, female 11 to 14, female 15 to 18. All modeling steps described below are applied identically within each subgroup and then replicated across the two country samples to support comparative analysis.

Additional Variable

Weekly usage intensity (Week_Use) was measured through self-reported frequency of ICT device use on a 5-point scale. This variable serves as a behavioral correlate for supplementary analysis. Missing values on Week_Use were minimal ($< 3\%$ across subgroups). For comparability across subgroups, Week_Use was z-standardized within the pooled sample.

2.3.2 Layer 1: Random Forest Discovery with Out of Fold SHAP Explanations

We use Random Forest regression as the primary discovery model because it can represent nonlinear relationships and high order interactions without requiring the analyst to pre specify functional forms. Random Forests also tend to be robust to noisy predictors and correlated features, which is important in multi dimensional constructs such as digital maturity (Breiman, 2001). Random Forests have also been discussed as useful tools for empirical social science and econometrics when the goal is flexible approximation rather than strict parametric inference (Mullainathan and Spiess, 2017).

A significant methodological challenge in machine learning-based interpretability is the risk of 'circular validation', where explanations are generated using the same data employed for model training. Such an approach can artificially inflate the perceived structure of SHAP (SHapley Additive exPlanations) values, effectively confounding true underlying signals with overfit idiosyncrasies. To ensure the robustness of our feature attributions and mitigate optimistic bias, we implement an out-of-fold (OOF) procedure for calculating both predictions and SHAP values.

The procedure follows a structured cross-fitting framework:

1. **Data Partitioning:** Each subgroup dataset is partitioned into K exhaustive and mutually exclusive folds.

2. **Iterative Training and Attribution:** For each fold $k \in \{1, \ldots, K\}$:
   - The Random Forest model is trained on the remaining $K - 1$ folds.
   - Point estimates for the target variable (age) are generated for the held-out fold k.
   - SHAP values are computed for the observations in fold k using the model instance that was trained entirely without those specific observations.

3. **Aggregation:** The held-out predictions and SHAP attributions are concatenated to form a complete, out-of-sample dataset for downstream analysis.

This approach is grounded in the principle of sample splitting and cross-fitting, a technique used to maintain the validity of downstream inference when utilizing highly flexible functional forms (Chernozhukov et al., 2018). By decoupling the estimation of the model from the calculation of feature importance, we align our work with established best practices in model selection that necessitate strict separation between training and validation phases to reduce overfitting-induced bias (Varma and Simon, 2006).

2.3.3 SHAP: Why and How it is Computed

We use SHAP (SHapley Additive exPlanations) to attribute each individual prediction to contributions from each feature. SHAP is grounded in cooperative game theory. The Shapley value provides a principled allocation of a total outcome to contributors under axioms such as local accuracy and consistency (Lundberg and Lee, 2017). For tree based models, TreeSHAP provides an efficient exact computation of SHAP values and supports aggregation from local explanations to global summaries (Lundberg et al., 2020). We summarize SHAP results in two complementary ways:

- **Beeswarm (Summary) plots**. These visualize the distribution of SHAP values for each feature, showing both typical direction and heterogeneity across individuals. Each point is one individual. The x axis is the SHAP contribution in predicted years. Color encodes the standardized feature value (low to high).

- **Global Importance via Mean Absolute SHAP**. For each feature j, we compute mean absolute SHAP across individuals within subgroup as a magnitude based importance measure. Importantly, mean absolute SHAP reflects contribution magnitude, not direction.

To avoid over interpreting small differences in importance ranks, we compute bootstrap confidence intervals for mean absolute SHAP. The bootstrap is a standard resampling approach for approximating sampling variability of statistics without strong parametric assumptions. In each subgroup we resample individuals with replacement, recompute mean absolute SHAP, and then use percentile intervals.

Layer 1 Functional Interpretation: PDP, ALE, and Local Slope

SHAP provides additive attributions conditional on the model, but scholars often prefer an interpretable 'shape' of the relationship between a predictor and the model output. We therefore complement SHAP with effect function visualizations.

Partial Dependence Plots (PDPs) visualize the average predicted outcome as one feature varies, marginalizing over the joint distribution of the other features (Friedman, 2001). Individual Conditional Expectation (ICE) plots extend PDPs by plotting individual level curves, revealing heterogeneity masked by the average (Goldstein et al., 2015). However, PDPs can be misleading when predictors are correlated because they evaluate the model at feature combinations that may be rare or unrealistic in the observed data. Therefore we rely on Accumulated Local Effects (ALE) plots that address the correlation problem by computing local differences in predictions over small intervals of a feature and then accumulating them. Because ALE integrates over the conditional distribution of the feature given the others, it tends to be more reliable than PDP when predictors are dependent (Apley and Zhu, 2020).All plots are generated on the standardized predictor scale so that x axes are comparable across features and across subgroups.

2.3.4 Layer 2: Confirmatory Spline Regression for Inference on Top SHAP Features

While Layer 1 identifies the most salient predictors through high-capacity machine learning, it does not provide the formal inferential framework required to assess statistical significance or quantify uncertainty in a way that is traditional in the social sciences. Consequently, we implement a second layer using semiparametric spline regression. This layer asks a stricter question: Given the features highlighted by the discovery model, do we observe statistically reliable, interpretable, and non-linear associations within the empirical support of each subgroup?

Spline Specification and Functional Flexibility: For each subgroup, we select the top k features identified by the mean absolute SHAP values in Layer 1. We then estimate a series of additive models where the relationship between each predictor and chronological age is modeled using a B-spline basis. The model is specified as:

$$\mathrm{Age}i = \alpha + \sum j = 1^k f_j(x_{ij}) + \epsilon_i \tag{2.1}$$

where each f_j is a non-parametric function represented by a cubic B-spline basis. B-splines are selected for their local control and numerical stability, allowing for the approximation of complex, non-linear functional forms without the global oscillations common in high-order polynomials (Virtanen et al., 2020). By using a cubic basis, we ensure that the estimated curves are twice-continuously differentiable, providing the 'smoothness' necessary for meaningful interpretation of marginal effects (Hastie, 2017). To ensure the validity of our estimates, we evaluate and plot marginal effect curves exclusively within the empirical support of the standardized predictors, thereby avoiding the risks associated with extrapolation beyond observed data ranges.

Joint Wald Tests for Feature Significance: In a spline framework, a single predictor x_j is represented by a vector of basis coefficients. Consequently, traditional t-tests on individual coefficients are insufficient to determine a feature's overall impact. To address this, we conduct a joint Wald test for each feature's spline block. We test the null hypothesis:

$$H_0 : \beta_{j,1} = \beta_{j,2} = \cdots = \beta_{j,m} = 0 \tag{2.2}$$

where m is the number of basis functions for feature j. This test determines whether the feature contributes any systematic association with age after accounting for other variables in the model. The Wald test serves as a robust large-sample diagnostic for restrictions on parameter vectors (Arend and Schäfer, 2019).

Inference and Multiple Testing Correction: A common challenge in subgroup analysis is the presence of heteroskedasticity, where the variance of the error term ϵ_i may correlate with the predictors. To ensure our inferences are not biased by such variance structures, we employ heteroskedasticity-consistent covariance estimates, specifically the HC3 variant. The HC3 estimator is preferred over earlier versions (like HC0 or HC1) because it incorporates a jackknife-like scaling factor that provides superior finite-sample performance and more conservative confidence intervals (Long and Ervin, 2000).Furthermore, because we conduct multiple joint tests across various features within each subgroup, the probability of Type I errors (false positives) increases. We address this by adjusting our p-values using the Benjamini-Hochberg (BH) procedure to control the False Discovery Rate (FDR). Unlike the overly conservative Bonferroni correction, the BH procedure is appropriate for discovery-confirmatory pipelines where the objective is to maintain power while controlling the expected proportion of false rejections (Weber, 2014).

2.4 Results

We present findings following the two-layer analytic structure outlined in the methodology, organized around the theoretical expectations developed in Section 2. The theoretical framework proposed that digital maturity development follows a phase-structured pattern where Digital Literacy operates as a foundational capability in early adolescence, self-regulatory dimensions become more salient in later adolescence, and the binding constraints differ by gender—with emotional regulation particularly important for older females due to affordance-vulnerability mismatch. We also examine within-age heterogeneity to assess whether chronological age adequately proxies developmental status. Layer 1 employs Random Forest models with out-of-fold SHAP to discover which dimensions associate most strongly with developmental stage. Layer 2 subjects these discoveries to formal statistical testing using semi-parametric spline regression with joint Wald tests. We conclude with predicted age profile analysis examining within-age heterogeneity.

2.4.1 Layer 1: Discovery Phase

The Austrian (n=1,721) and the German sample (n=1,297) reveal distinct developmental patterns across four subgroups and provides critical evidence about which patterns generalize across contexts. Sample statistics are provided in Table 2.2 for all the four groups. Layer 1 discovery results are visualized through SHAP beeswarm plots showing all dimensions ranked by importance (Figures 2.1 and 2.6) and ALE (Accumulated Local Effects) curves revealing functional forms for the 4 most influential dimensions in each subgroup. All predictors are z-standardized, enabling direct comparability across features and subgroups.

Table 2.2: Summary statistics: sample size and average age by group and country

	Austria		Germany	
Group	**N**	**Average age**	**N**	**Average age**
Males (11–14)	517	13.29	343	13.59
Females (11–14)	477	13.35	358	13.41
Males (15–18)	368	16.51	295	16.40
Females (15–18)	351	16.55	298	16.36

Austria: Developmental Differentiation Across Age and Gender

Males Aged 11–14. The SHAP beeswarm plot (Figure 2.1 reveals a clear hierarchy among digital maturity dimensions. Digital Literacy ranks first by substantial margin, with mean absolute SHAP values approximately three times larger than secondary dimensions. The beeswarm shows tight clustering of positive SHAP values for high-literacy adolescents (red points) and negative values for low-literacy adolescents (blue points), with minimal vertical spread indicating strong individual consistency. Autonomy Within Digital Context ranks second, Individual Growth third, and Autonomy of Choice fourth. All these dimensions show modest positive SHAP distributions with mean absolute values approximately 30–40% of literacy magnitude. Emotional regulation and prosocial dimensions cluster in the bottom half with SHAP values near zero.

Examining functional forms through ALE curves in Figure 2.2, Digital Literacy exhibits a monotonic positive pattern, rising steadily from negative values at low z-scores to positive values around z=+0.5, then plateauing at higher values. This pattern indicates that literacy gains contribute positively to predicted age up to a threshold, beyond which additional literacy shows diminishing returns. Autonomy Within Digital Context shows a weaker S-shaped positive pattern with modest overall effect magnitude. Individual Growth displays a slight inverted-U form, rising to peak around z=+1.5 before declining at extreme high values. Autonomy of Choice exhibits a U-shaped pattern, starting elevated at low values, dipping in the mid-range, then rising again at high values, suggesting non-monotonic developmental associations.

Males Aged 15–18. Digital Literacy maintains first position with even larger mean absolute SHAP values than younger males in Figure 2.1. However, the ranking composition shifts as we observe Regulation of Aggressive Impulses rises to second position, a dimension ranking in the bottom half for younger males. Support Seeking Behavior ranks third, and Autonomy of Choice fourth. This reordering of behavioral regulation and prosocial help-seeking displacing autonomy dimensions, signals a developmental transition during ages 15–18.

The ALE curves reveal distinctive functional forms in Figure 2.3. Digital Literacy shows an unusual step-down pattern, a very high positive values at low z-scores (around +0.07), dropping sharply near z=-1.5 to approximately -0.01, then remaining flat across the rest of

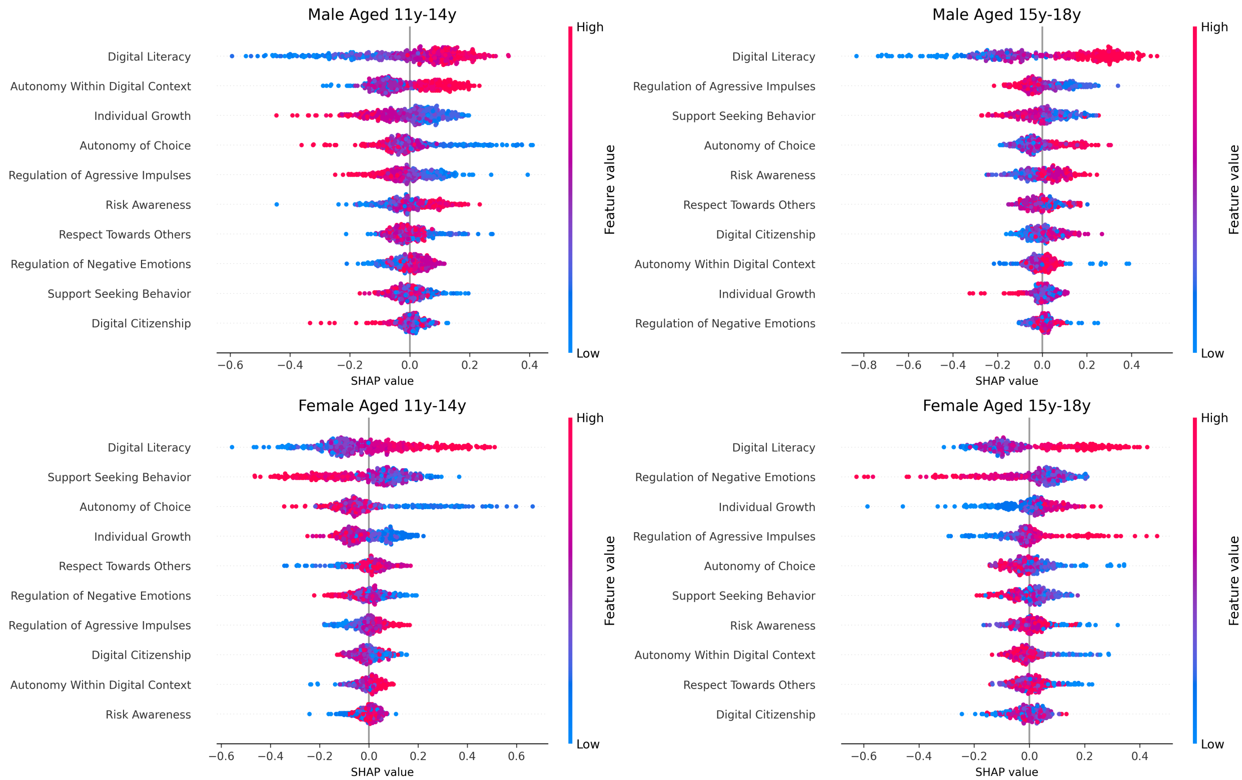

Figure 2.1: SHAP Beeswarm Plots for Austria

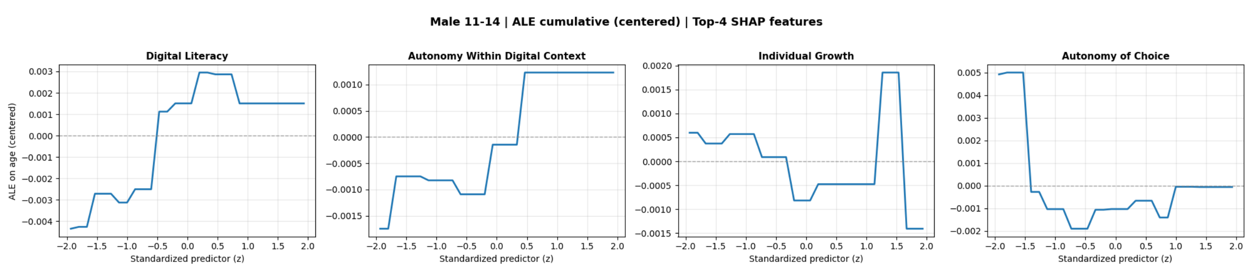

Figure 2.2: Austria Male 11-14 ALE Plots for Top 4 Dimensions

the range. This suggests that low literacy is particularly diagnostic of younger developmental status, while moderate-to-high literacy levels do not further differentiate age. Regulation of Aggressive Impulses similarly shows a step-down pattern, elevated at low values, dropping at mid-range, then flat. Support Seeking Behavior exhibits a positive monotonic pattern, rising steadily from approximately -0.004 to +0.01 across the z-score range. Autonomy of Choice shows positive trend with elevation at high values.

Females Aged 11–14. Digital Literacy dominates again with patterns visually indistinguishable from younger males in this group. The critical gender divergence appears in secondary positions, Support Seeking Behavior shows clear positive SHAP associations which is a pattern absent from younger male rankings. Autonomy of Choice ranks third and Individual Growth fourth. Emotional regulation dimensions remain lower-ranked with SHAP values near zero.

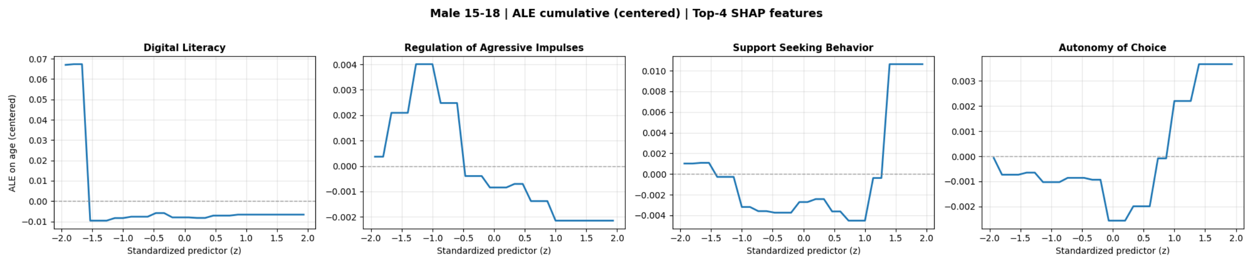

Figure 2.3: Austria Male 15-18 ALE Plots for Top 4 Dimensions

The ALE curves reveal gender-specific functional forms in Figure 2.4. Digital Literacy shows a positive monotonic pattern, rising from approximately -0.006 at low z-scores to +0.002 at high values, with steeper gains at lower values. Support Seeking Behavior exhibits an inverted-U pattern, rising to a peak around z=0, then declining to negative values at high z-scores. This suggests that moderate support-seeking is associated with older predicted age, but very high levels may indicate developmental immaturity. Autonomy of Choice similarly shows an inverted-U form with mid-range peak. Individual Growth displays a step-down pattern, starting high at low z-scores.

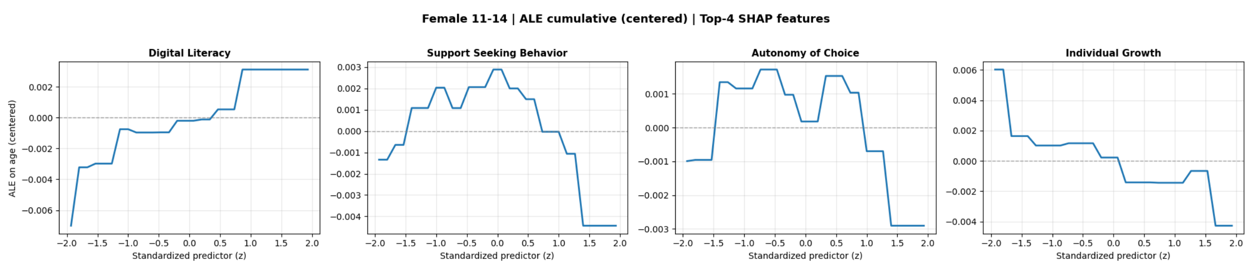

Figure 2.4: Austria Female 11-14 ALE Plots for Top 4 Dimensions

Females Aged 15–18. The SHAP beeswarm for this Austrian group reveals Regulation of Negative Emotions ranks first, displacing Digital Literacy from its universal position of dominance. The beeswarm shows remarkably tight clustering as high-regulation adolescents contribute large positive SHAP values ranging from +0.8 to +1.8 predicted years, while low-regulation adolescents contribute large negative values. Digital Literacy ranks second, Regulation of Aggressive Impulses third, and Individual Growth fourth.

The ALE curves Figure 2.5 explain why emotional regulation dominates. Regulation of Negative Emotions shows a step-down pattern as high positive values (around +0.003) at low z-scores, dropping to negative values (around -0.003) at mid-range, then remaining flat. This indicates that low emotional regulation is particularly diagnostic of younger developmental status among older females. Digital Literacy shows a positive S-shaped pattern, rising from approximately -0.003 to +0.003. Individual Growth exhibits positive monotonic rise. Most notably, Regulation of Aggressive Impulses shows a strong positive monotonic pattern, rising from approximately -0.008 to +0.006, the steepest gradient among the four dimensions, suggesting that impulse regulation differentiates developmental age continuously across the full range.

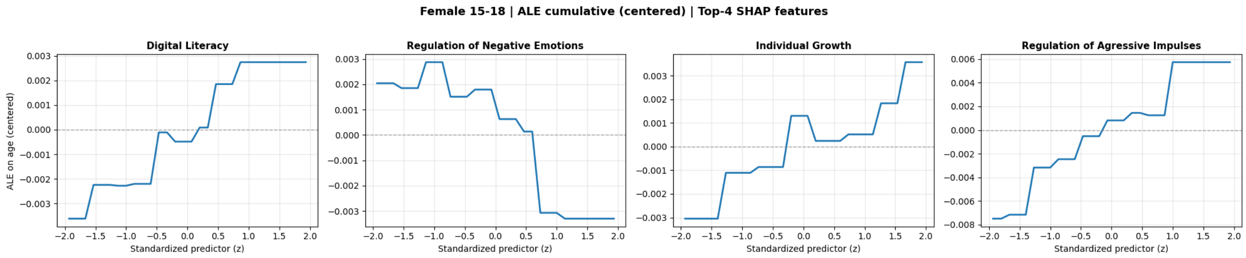

Figure 2.5: Austria Female 15-18 ALE Plots for Top 4 Dimensions

Germany: Cross-National Patterns and Divergences

Males Aged 11–14. Digital Literacy dominates with patterns similar to Austrian younger males with highest mean absolute SHAP by large margin, tight clustering of strongly positive values for high literacy as seen in Figure 2.6. Respect Towards Others ranks second which is a prosocial dimension absent from Austrian younger male top four. Support Seeking Behavior ranks third and Risk Awareness fourth. These secondary dimensions show notably lower mean absolute SHAP values, approximately 30–35% of literacy magnitude compared to 40–50% in Austria suggesting weaker secondary signals.

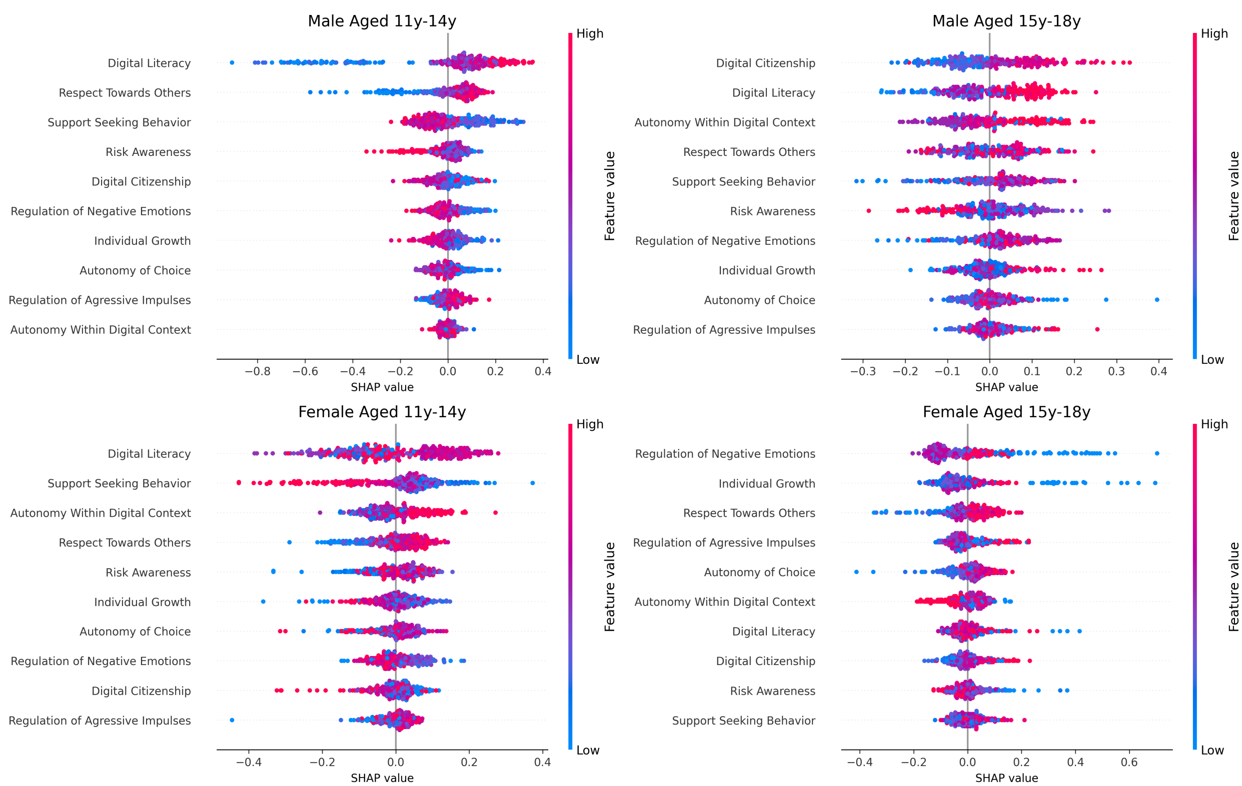

Figure 2.6: SHAP Beeswarm Plots for Germany

The ALE curves reveal both convergence and divergence with Austria. Digital Literacy for this German group, as seen in Figure 2.7 shows an S-shaped positive pattern, dipping slightly around $z = -0.5$, then rising sharply to approximately $+0.0035$, and plateauing at high values. Respect Towards Others exhibits a similar S-shaped positive pattern, rising from approximately -0.005 to $+0.004$ around $z = 0$, then plateauing. Support Seeking Behavior shows an inverted-U pattern, peaking around $z = -0.5$, then declining to negative values at high z-scores indicating that moderate support-seeking associates with older predicted age but excessive levels do not. Risk Awareness displays a step-down pattern, starting positive and dropping sharply at $z \approx +1$.

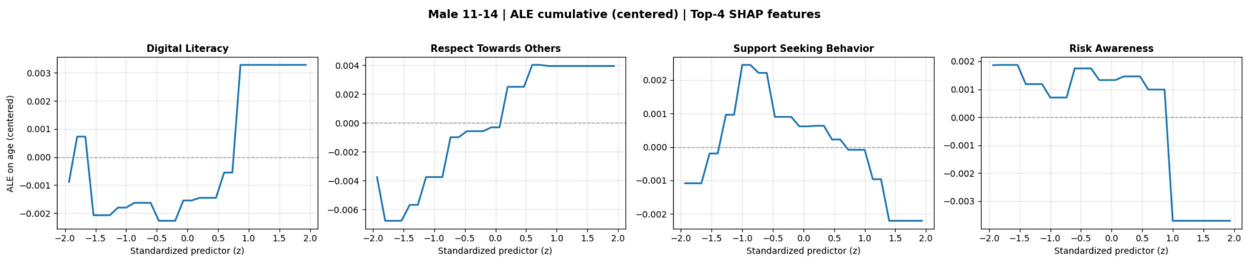

Figure 2.7: Germany Male 11-14 ALE Plots for Top 4 Dimensions

Males Aged 15–18. German older males present a unique pattern as Digital Citizenship ranks first, the only instance across both countries where Digital Literacy does not occupy the top position. Digital Literacy ranks second, Autonomy Within Digital Context third, and Respect Towards Others fourth. The prominence of civic-prosocial dimensions contrasts with Austrian older males where behavioral regulation appeared more prominently.

ALE curves reveal functional complexity in Figure 2.8. Digital Citizenship exhibits an S-shaped positive pattern which is flat at approximately -0.002 until $z \approx 0$, then rising sharply to $+0.005$ and plateauing. This threshold pattern suggests civic awareness becomes developmentally differentiating only at moderate-to-high levels. Digital Literacy shows a similar S-shaped positive pattern, rising from approximately -0.004 to $+0.003$. Autonomy Within Digital Context displays a step-down pattern with partial recovery, starting high around $+0.006$, dropping to -0.002, then recovering slightly. Respect Towards Others shows a U-shaped pattern, starting elevated, dropping to approximately -0.004 at mid-range, then recovering.

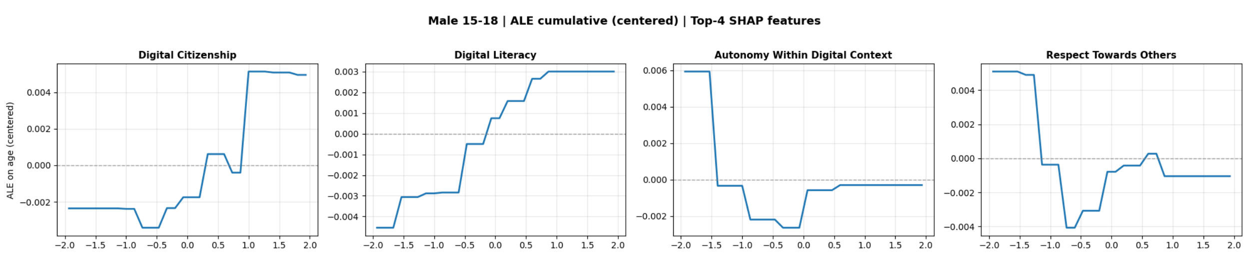

Figure 2.8: Germany Male 15-18 ALE Plots for Top 4 Dimensions

Females Aged 11–14. Digital Literacy ranks first but with notably lower mean absolute SHAP magnitude than Austrian younger females or German younger males. Support Seeking Behavior ranks second, Autonomy Within Digital Context third, and Respect Towards Others fourth. Across all four top dimensions, SHAP patterns are notably weaker and more diffuse than corresponding Austrian subgroup.

The ALE curves reveal patterns that help explain the weaker Layer 2 confirmations in Figure 2.9. Digital Literacy shows a step-down pattern, starting high at approximately +0.006, then dropping sharply and remaining flat or negative across most of the range. This contrasts with the positive monotonic pattern observed in Austrian younger females. Support Seeking Behavior similarly shows a step-down pattern, declining steadily from +0.006 to −0.008 indicating that low support-seeking associates with older predicted age, the opposite direction from Austria. Autonomy Within Digital Context shows a positive rise followed by plateau. Respect Towards Others exhibits positive monotonic pattern, rising from approximately −0.005 to +0.002.

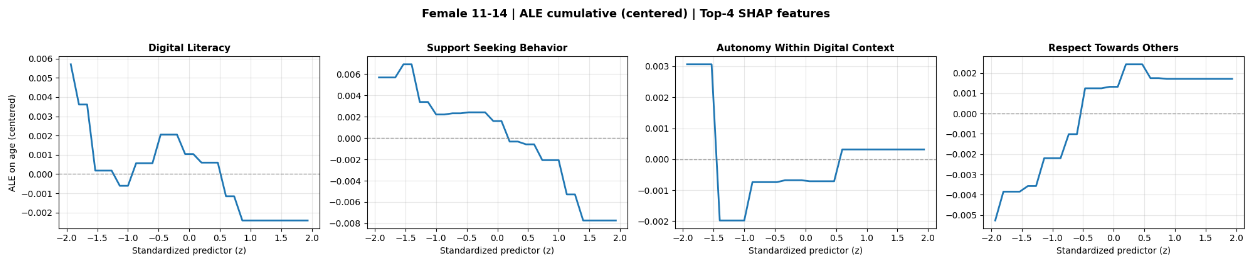

Figure 2.9: Germany Female 11-14 ALE Plots for Top 4 Dimensions

Females Aged 15–18. The SHAP beeswarm shows Regulation of Negative Emotions ranking first, paralleling the Austrian pattern. Individual Growth ranks second, Respect Towards Others third, and Regulation of Aggressive Impulses fourth. The beeswarm pattern shows clustering of positive SHAP values for high-regulation individuals and negative values for low-regulation individuals. Digital Literacy ranks lower and exhibits more dispersed SHAP distributions, paralleling the Austrian older female pattern.

The ALE curves reveal a critical cross-national divergence in functional form in Figure 2.10. Regulation of Negative Emotions shows a U-shaped pattern that then rises, dropping from approximately +0.0005 to −0.004 around $z = -0.5$, then rising sharply to +0.004 at high values. This contrasts with the Austrian step-down pattern, suggesting different developmental dynamics. In Germany, high emotional regulation associates with older predicted age (positive relationship at high values), while in Austria, low emotional regulation is particularly diagnostic of younger status (negative relationship concentrated at low values). Individual Growth shows an S-shaped positive pattern. Respect Towards Others exhibits a step-down then recovery pattern. Regulation of Aggressive Impulses shows a step-down pattern starting elevated around +0.004, then dropping and remaining flat.

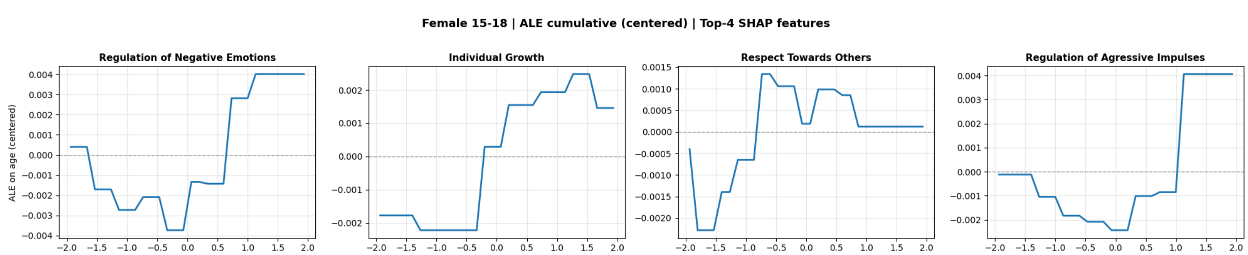

Figure 2.10: Germany Female 15-18 ALE Plots for Top 4 Dimensions

2.4.2 Layer 2: Confirmatory Statistical Testing

Layer 2 subjects the dimensions identified in Layer 1 to formal hypothesis testing, asking whether the patterns discovered through flexible machine learning withstand rigorous statistical scrutiny. This step is essential for moving from pattern discovery to defensible theoretical claims. We use semiparametric B-spline regression with joint Wald tests, HC3 robust standard errors, and Benjamini-Hochberg FDR correction. Table 2.3 presents complete confirmatory test results.

Table 2.3: Layer 2 Confirmatory Test Results

Austria				**Germany**		
Dimension	$\chi^2_{(4)}$	p_{FDR}	Subgroup	Dimension	$\chi^2_{(4)}$	p_{FDR}
Digital Literacy	20.08	.002✓	M 11–14	Digital Literacy	35.68	<.001✓
Autonomy of Choice	13.03	.022✓	M 11–14	Respect Towards Others	12.02	.034✓
Autonomy Within Digital Context	10.88	.034✓	M 11–14	Support Seeking Behavior	9.69	.061†
Individual Growth	10.41	.034✓	M 11–14	Risk Awareness	7.63	.106
Digital Literacy	35.97	<.001✓	M 15–18	Digital Literacy	3.31	.791
Regulation of Aggressive Impulses	8.77	.134†	M 15–18	Autonomy Within Digital Context	3.10	.791
Support Seeking Behavior	5.65	.258	M 15–18	Respect Towards Others	2.79	.791
Autonomy of Choice	5.30	.258	M 15–18	Digital Citizenship	1.62	.805
Digital Literacy	27.14	<.001✓	F 11–14	Digital Literacy	15.20	.017✓
Autonomy of Choice	14.13	.014✓	F 11–14	Support Seeking Behavior	8.90	.127
Support Seeking Behavior	12.51	.019✓	F 11–14	Respect Towards Others	3.80	.578
Individual Growth	5.04	.284	F 11–14	Autonomy Within Digital Context	1.21	.876
Regulation of Negative Emotions	18.31	.004✓	F 15–18	Regulation of Negative Emotions	11.29	.094†
Digital Literacy	14.09	.014✓	F 15–18	Respect Towards Others	6.67	.210
Regulation of Aggressive Impulses	9.48	.067†	F 15–18	Individual Growth	6.61	.210
Individual Growth	4.07	.397	F 15–18	Regulation of Aggressive Impulses	1.07	.899

Note: ✓ = $p_{\text{FDR}} < .05$; † = approaching significance ($p_{\text{FDR}} < .15$).

Austria: Strong Confirmation Patterns and Marginal Effects

Austria exhibits strong confirmation across 10 of 16 tested dimension-subgroup combinations (62.5%). Digital Literacy confirms in all four subgroups, establishing it as a robust developmental marker across age and gender. An additional 2 dimensions approach significance at relaxed thresholds ($p_{\text{FDR}} < 0.15$), suggesting emerging but not yet robust associations. For younger males, we observe a comprehensive multidimensional development as this subgroup confirms all four dimensions. Digital Literacy ($\chi^2 = 20.08$, $p_{\text{FDR}} = 0.002$) shows the strongest effect. Autonomy of Choice ($\chi^2 = 13.03$, $p_{\text{FDR}} = 0.022$), Autonomy Within Digital Context ($\chi^2 = 10.88$, $p_{\text{FDR}} = 0.034$), and Individual Growth ($\chi^2 = 10.41$, $p_{\text{FDR}} = 0.034$) also achieve significance. This comprehensive confirmation indicates early male adolescence represents a period of broad developmental differentiation, technical competencies, autonomous decision-making, and personal growth all systematically relate to chronological age.

For the older males, we see concentrated literacy effect as only Digital Literacy confirms ($\chi^2 = 35.97$, $p_{\text{FDR}} < 0.001$) the strongest single chi-square value in Austria. Regulation of

Aggressive Impulses approaches significance at relaxed thresholds ($p = 0.067$, $p_{\mathrm{FDR}} = 0.134$). Support Seeking Behavior and Autonomy of Choice fail to reach significance. The transition from four confirmations in younger males to one in older males suggests developmental convergence, by ages 15–18, Austrian males have achieved sufficient homogeneity in autonomy and growth dimensions that these no longer discriminate age, while literacy continues differentiating.

For younger females we find prosocial emergence as three dimensions confirm, Digital Literacy ($\chi^2 = 27.14$, $p_{\mathrm{FDR}} < 0.001$), Autonomy of Choice ($\chi^2 = 14.13$, $p_{\mathrm{FDR}} = 0.014$), and Support Seeking Behavior ($\chi^2 = 12.51$, $p_{\mathrm{FDR}} = 0.019$). The confirmation of Support Seeking Behavior for females but not males represents statistically validated gender differentiation, prosocial help-seeking emerges as developmentally salient earlier in female trajectories. For older females, emotional regulation emergence as two dimensions confirm, including Regulation of Negative Emotions ($\chi^2 = 18.31$, $p_{\mathrm{FDR}} = 0.004$) alongside Digital Literacy ($\chi^2 = 14.09$, $p_{\mathrm{FDR}} = 0.014$). Regulation of Aggressive Impulses approaches significance ($p = 0.050$, $p_{\mathrm{FDR}} = 0.067$). The Layer 1 ALE analysis revealed that Regulation of Negative Emotions displays a step-down functional form (high values at low z-scores dropping to lower values at mid-range) indicating that low emotional regulation is particularly diagnostic of younger developmental status. This confirmation of emotional regulation, establishes affective self-management as a defining developmental frontier for older female adolescents.

Germany: Concentrated Confirmation with Notable Null Findings

Germany shows confirmation for only 3 of 16 combinations (18.75%), with an additional 2 dimensions approaching significance at relaxed thresholds. This substantially lower confirmation rate indicates either more concentrated developmental signals or greater measurement challenges in the German context. For younger males we observe strong literacy with prosocial emergence as two dimensions confirm here, Digital Literacy ($\chi^2 = 35.68$, $p_{\mathrm{FDR}} < 0.001$) and Respect Towards Others ($\chi^2 = 12.02$, $p_{\mathrm{FDR}} = 0.034$). Digital Literacy represents the strongest single finding across both countries. Support Seeking Behavior approaches significance ($p = 0.046$, $p_{\mathrm{FDR}} = 0.061$). Risk Awareness fails to reach significance. The confirmation of prosocial Respect Towards Others in Germany but not Austria represents cross-national divergence in early male developmental patterns. For older males there is a complete null finding as this subgroup shows no confirmations. All four dimensions fail with remarkably weak chi-square values: Digital Citizenship ($\chi^2 = 1.62$, $p_{\mathrm{FDR}} = 0.805$), Digital Literacy ($\chi^2 = 3.31$, $p_{\mathrm{FDR}} = 0.791$), Autonomy Within Digital Context ($\chi^2 = 3.10$, $p_{\mathrm{FDR}} = 0.791$), and Respect Towards Others ($\chi^2 = 2.79$, $p_{\mathrm{FDR}} = 0.791$). This comprehensive null pattern indicates that by ages 15–18, German males may have achieved developmental convergence across all measured dimensions (within the dataset), such that none discriminates chronological age. Several interpretations warrant consideration like German older males may exhibit unusually low developmental heterogeneity or cultural or educational factors may produce more uniform trajectories or DIMI dimensions may not capture the most salient developmental processes for this demographic.

For younger females in Germany we observe weak confirmation as only Digital Literacy confirms ($\chi^2 = 15.20$, $p_{\mathrm{FDR}} = 0.017$). Support Seeking Behavior approaches nominal significance ($p = 0.064$) but fails FDR correction ($p_{\mathrm{FDR}} = 0.127$). Respect Towards Others and

Autonomy Within Digital Context fail to reach significance. The contrast with Austrian younger females (three confirmations) suggests either greater developmental homogeneity in the German sample or that DIMI dimensions do not effectively capture differentiation processes for German younger females. For older females we see partial emotional regulation important as no dimensions achieve FDR-corrected significance, though Regulation of Negative Emotions achieves nominal significance ($p = 0.023$) and approaches correction at relaxed thresholds ($p_{\text{FDR}} = 0.094$). Respect Towards Others and Individual Growth show modest positive trends ($p \approx 0.155$–0.158) but fail confirmation.

Three major patterns emerge from cross-national comparison of the above results. The first one is about the importance of Digital Literacy in Early Adolescence. Digital Literacy confirms in 5 of 8 subgroups with the two strongest chi-square values in the study (Austria M15–18: $\chi^2 = 35.97$; Germany M11–14: $\chi^2 = 35.68$). Layer 1 ALE curves show consistent functional forms where confirmed: persistently positive derivatives around 0.15–0.25 per SD with minimal variation across the predictor range, indicating near-linear continuous development. However, literacy weakens in older subgroups (German males 15–18 show no confirmation ($\chi^2 = 3.31$), and older females show lower chi-square values than younger subgroups) suggesting developmental convergence where populations achieve baseline competence and literacy ceases to discriminate age.Second, we observe gender-differentiated prosocial development. Prosocial dimensions exhibit systematic gender patterns. Support Seeking Behavior confirms only for Austrian females aged 11–14 ($\chi^2 = 12.51$, $p_{\text{FDR}} = 0.019$) and approaches significance for German younger males ($p_{\text{FDR}} = 0.061$), absent from Austrian male subgroups entirely. Respect Towards Others confirms for German younger males ($\chi^2 = 12.02$, $p_{\text{FDR}} = 0.034$) but not Austrian younger males, representing cross-national divergence in male prosocial development. The ALE curves confirm prosocial dimensions consistently show curvilinearity with mid-range emphasis, flat at low values, steeper in mid-range, moderating at high values, indicating threshold-like processes rather than linear accumulation.

Lastly, emotional regulation is female-specific frontier in late adolescence. The most theoretically significant finding is emotional regulation's unique salience for older females. In Austria, Regulation of Negative Emotions confirms strongly ($\chi^2 = 18.31$, $p_{\text{FDR}} = 0.004$) and Regulation of Aggressive Impulses approaches significance ($p_{\text{FDR}} = 0.067$). In Germany, Regulation of Negative Emotions achieves nominal significance ($p = 0.023$, $p_{\text{FDR}} = 0.094$). Critically, no male subgroup in either country shows emotional regulation among top-ranked dimensions or approaching significance. The Layer 1 ALE curves across both countries show distinctive inverted-U patterns with accelerated mid-range development—slopes reaching maximum around the distribution center. This convergent evidence (consistent functional forms, and high specificity only older females) suggests emotional self-management represents a genuine developmental phenomenon for older female adolescents, with cross-national replication limited by statistical power rather than absence of the underlying process.

Three patterns emerge from cross-national comparison that directly address the theoretical framework. First, Digital Literacy's foundational role receives robust support—it confirms in 5 of 8 subgroups with the two strongest chi-square values in the study, and Layer 1 ALE curves show consistent positive functional forms where confirmed. Second, gender-differentiated prosocial development receives partial support—Support Seeking Behavior confirms for Austrian younger females but not males, and Respect Towards Others confirms for German younger males but not Austrian younger males, indicating that prosocial dimensions follow gender- and context-specific trajectories. Third, and most significantly, emotional regulation's unique salience for older females receives cross-national validation. In Austria, Regulation of Negative Emotions confirms strongly ($\chi^2 = 18.31$, p_FDR = 0.004); in Germany, it achieves nominal significance (p = 0.023, p_FDR = 0.094). No male subgroup in either country shows emotional regulation among top-ranked dimensions or approaching significance. This convergent evidence—replicated across independent national samples—provides strong support for the affordance-vulnerability mismatch.

2.4.3 Predicted Age Profiles: Developmental Configurations and Within-Age Heterogeneity

The preceding analyses establish which digital maturity dimensions associate with chronological age and with what functional forms. We now examine a complementary question with direct policy relevance: to what extent do adolescents of identical chronological age differ in their developmental profiles, and can digital maturity configurations systematically account for this heterogeneity? This analysis proceeds in a three step chain. First, we document the existence of within-age dispersion in predicted developmental age. Second, we demonstrate that this dispersion is systematically related to maturity configurations amd third, we illustrate through individual case decomposition how multidimensional profiles translate into developmental advancement or delay.

Step 1: Documenting Within-Age Prediction Heterogeneity

We operationalize within-age heterogeneity as dispersion in predicted developmental age among adolescents who share the same chronological age. Rather than relying on aggregate age-band comparisons, we examine exact-age distributions to rule out compositional effects. Conceptually, if digital maturity captures meaningful developmental variance, then adolescents of the same chronological age should still exhibit non-trivial spread in predicted age, and maturity-based strata should separate within those exact ages. Figures 2.11 and 2.12 visualizes predicted age distributions at each exact chronological age, stratified by maturity tercile, for all four demographic subgroups across both countries.

Two descriptive regularities emerge from these figures. First, within-age dispersion is present across all subgroups in both countries. Even holding chronological age constant, the predicted age distributions retain meaningful spread, typically spanning 0.5 to 1.0 years within a single tercile and 1.0 to 1.5 years across all terciles combined at a given chronological age.

Second, separation between maturity strata varies substantially by subgroup, age, and national context, suggesting that within-age heterogeneity is structured rather than uniform. Among younger males (11–14) in Germany, tercile separation is modestly visible at ages 11–12, where the high-maturity tercile (green) shows slight upward displacement relative to the low-maturity tercile (blue); however, this separation diminishes at ages 13–14, where distributions converge. In Austria, younger males exhibit a different pattern: at age 11, the low-maturity tercile (blue) actually predicts higher developmental age than the high-maturity tercile (green), suggesting reversed or inconsistent tercile ordering. Among older

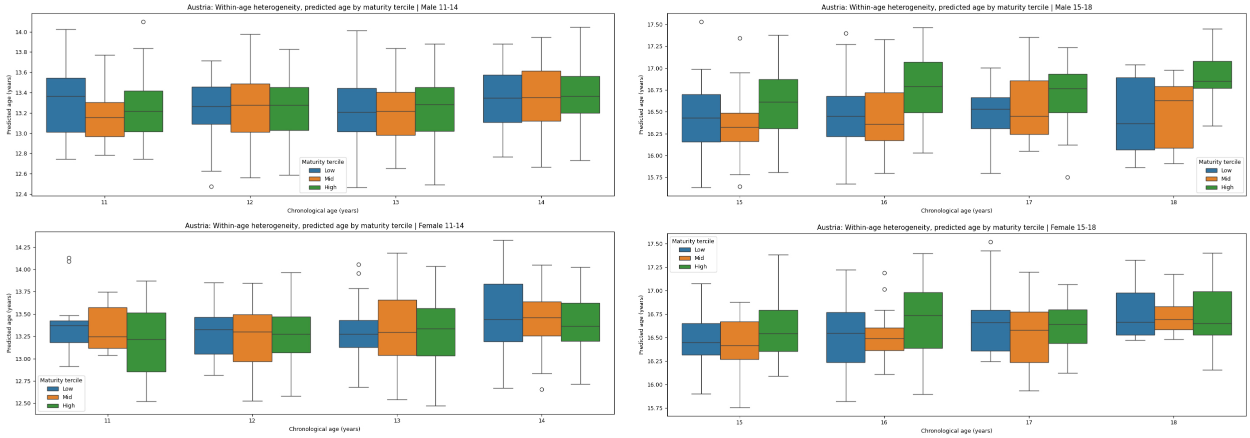

Figure 2.11: Austria Within-age heterogeneity in predicted developmental age by maturity tercile.

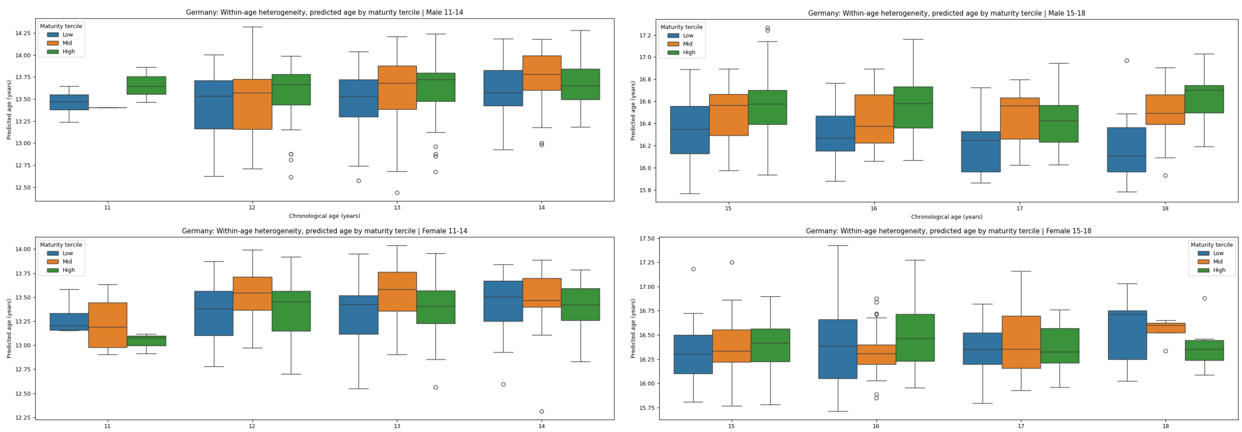

Figure 2.12: Germany Within-age heterogeneity in predicted developmental age by maturity tercile.

males (15–18) in both countries, tercile distributions overlap substantially, with no consistent separation between maturity strata, the green, orange, and blue boxes largely coincide at each chronological age. For females aged 11–14 in both countries, tercile ordering is inconsistent and sometimes reversed, with the high-maturity group predicting at or below the low-maturity group; this reversal is most pronounced in Germany at ages 11–12, where the high-maturity tercile sits visibly below the low-maturity tercile. Among older females (15–18), tercile separation remains minimal in both national contexts, with distributions heavily overlapping across all chronological ages.

From Table 2.4, which is a within age predictions gaps by chronological age, we also observe several patterns. For Male 15-18 in both countries, within-age gaps are consistently positive and substantial. Austrian 16-year-olds show a +0.30 year gap; German 16-year-olds show a +0.29 year gap; Austrian 18-year-olds show the largest gap at +0.46 years (approximately 5.5 months). These within-exact-age differences confirm that maturity-based separation is not an artifact of age-band aggregation. For Male 11-14, cross-national patterns

Table 2.4: Within-Age Prediction Gaps (High – Low Maturity Tercile) by Chronological Age

Gender	Age	Austria n Low	Austria n High	Austria Pred. Mean Low	Austria Pred. Mean High	Austria Gap (yrs)	Germany n Low	Germany n High	Germany Pred. Mean Low	Germany Pred. Mean High	Germany Gap (yrs)
Male	11	12	13	13.33	13.28	−0.05	4	3	13.46	13.66	+0.20
	12	59	52	13.26	13.25	−0.00	23	26	13.41	13.56	+0.15
	13	55	53	13.22	13.27	+0.05	44	30	13.48	13.60	+0.12
	14	47	55	13.33	13.38	+0.06	43	56	13.59	13.68	+0.09
	15	49	28	16.43	16.62	+0.19	47	42	16.32	16.56	+0.23
	16	40	49	16.47	16.78	+0.30	27	26	16.30	16.59	+0.29
	17	24	32	16.49	16.72	+0.24	16	18	16.21	16.41	+0.21
	18	10	14	16.43	16.89	+0.46	8	13	16.21	16.61	+0.40
Female	11	12	13	13.41	13.20	−0.21	4	3	13.29	13.04	−0.25
	12	47	55	13.28	13.25	−0.03	39	41	13.35	13.37	+0.02
	13	49	47	13.27	13.32	+0.04	38	43	13.32	13.38	+0.06
	14	51	44	13.49	13.38	−0.11	38	33	13.43	13.39	−0.04
	15	39	36	16.46	16.59	+0.14	41	39	16.33	16.39	+0.06
	16	42	41	16.55	16.67	+0.12	31	34	16.38	16.48	+0.09
	17	25	27	16.69	16.60	−0.09	18	17	16.35	16.35	−0.00
	18	11	13	16.77	16.79	+0.02	9	10	16.54	16.37	−0.17

Note: n Low/High = sample size for Low/High maturity tercile; Pred. Mean = mean predicted age; Gap = Pred. Mean (High) − Pred. Mean (Low). Positive gaps indicate high-maturity adolescents predict older than low-maturity peers of identical chronological age.

diverge. German younger males show consistently positive gaps (+0.09 to +0.20 years) at every age, while Austrian younger males fluctuate around zero with some negative values. For Female 11-14, gaps hover near zero and frequently turn negative in both countries (e.g., -0.21 years for Austrian 11-year-olds; -0.25 years for German 11-year-olds), confirming that the null finding for this subgroup is not an aggregation artifact. For Female 15-18, gaps show modest positive values at ages 15–16 in both countries (+0.12 to +0.14 years in Austria; +0.06 to +0.09 years in Germany) but attenuate or reverse at ages 17–18, suggesting that whatever developmental processes differentiate older females may concentrate at specific ages within the band rather than operating uniformly.

The key conclusion from Step 1 is methodological as dispersion exists within exact chronological age, the magnitude and direction of within-age gaps vary systematically by subgroup and national context, and this structured heterogeneity motivates a systematic maturity-based explanation tested in Step 2.

Step 2: Maturity Configurations Systematically Explain Prediction Variance

Having established that within-age dispersion exists, we examine whether digital maturity configurations systematically account for this variance. We construct an overall maturity profile index as the mean of all standardized (z-scored) maturity dimensions for each individual. Within each subgroup, defined by country, gender, and age band, we classify adolescents into Low, Mid, and High maturity terciles based on the 33rd and 66th percentiles of this index.

This within-subgroup classification isolates heterogeneity among same-age adolescents, preventing the trivially true observation that older adolescents have higher maturity on average from driving group definitions.

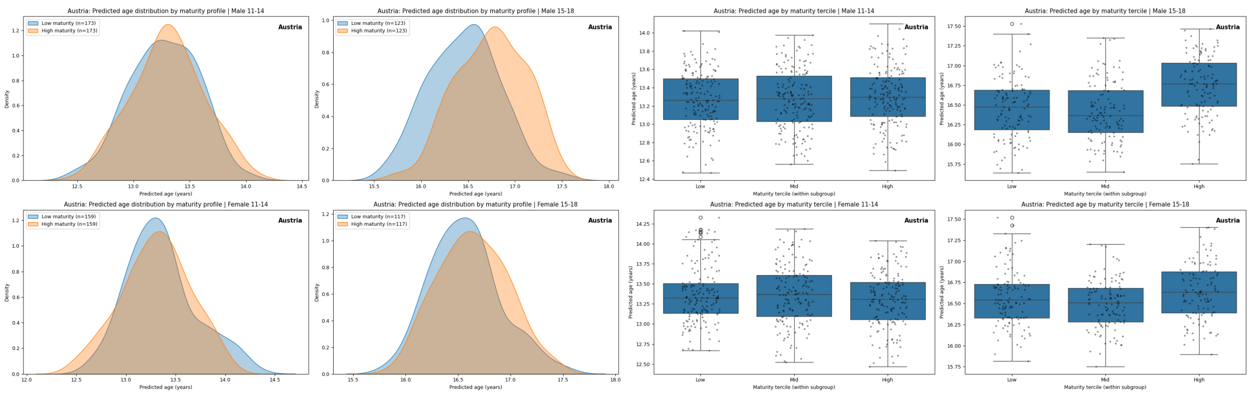

Figure 2.13: Austria Predicted Age by Maturity Tercile (Low, Mid, High) for Each Subgroup

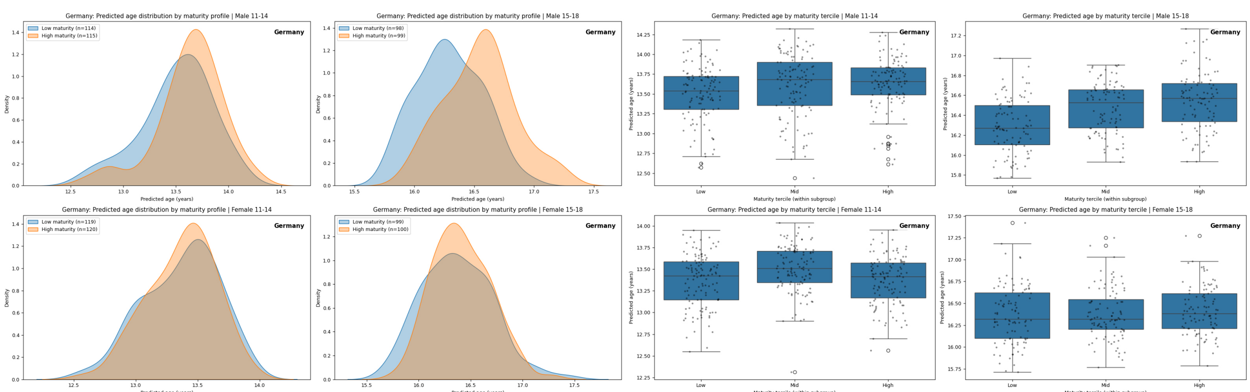

Figure 2.14: Germany Predicted Age by Maturity Tercile (Low, Mid, High) for Each Subgroup

The maturity-age relationship is most effectively assessed visually. Figures 2.13 and 2.14 presents density plots comparing low- versus high-maturity predicted age distributions, and boxplots showing the full tercile ordering with individual data points. Across both countries, Figures 2.13 and 2.14 reveal consistent cross-national patterns. For older males (15-18), predicted age distributions show clear rightward displacement as maturity increases. High-maturity males predict systematically older than low-maturity males within the same age band. The density plots reveal minimal overlap between low and high distributions, and the boxplots show clear upward progression across terciles. This pattern replicates across Austria and Germany, providing strong evidence that digital maturity configurations meaningfully differentiate developmental status among older male adolescents.

For younger females (11-14) in both countries, distributions overlap almost completely, and the boxplots show no consistent tercile ordering. In Austria, the high-maturity tercile actually shows slightly lower predicted age than the low-maturity tercile, indicating a reversed relationship. In Germany, the distributions are essentially superimposed, indicating negligible maturity-age association. For younger males (11-14), cross-national patterns diverge. German younger males show visible rightward displacement from Low to High tercile, while Austrian younger males show near-complete overlap. For older females (15-18), both countries show modest rightward displacement, weaker than older males but directionally consistent. Table 2.5 quantifies these patterns through effect size analysis, comparing High versus Low terciles within each subgroup.

Table 2.5: High vs. low maturity tercile comparison (cross-national)

	Austria					**Germany**				
Subgroup	**Low** *n*	**High** *n*	**Low** *Pred. age*	**High** *Pred. age*	**Cohen's d**	**Low** *n*	**High** *n*	**Low** *Pred. age*	**High** *Pred. age*	**Cohen's d**
M 11–14	173	173	13.27	13.30	0.10^{n}	114	115	13.51	13.63	0.38^{s}
M 15–18	123	123	16.46	16.74	0.77^{m}	98	99	16.29	16.55	0.90^{l}
F 11–14	159	159	13.36	13.30	-0.16^{n}	119	120	13.36	13.37	0.02^{n}
F 15–18	117	117	16.57	16.64	0.22^{s}	99	100	16.37	16.41	0.14^{n}

Note: n = sample size; Pred. Age = mean predicted age; d = Cohen's d (High − Low / pooled SD). Effect sizes: ^{n}negligible ($|d| < 0.20$); ^{s}small (0.20–0.50); ^{m}medium (0.50–0.80); ^{l}large (>0.80).

With the results from the above 2 steps and calculating Cohen's d for these groups, three patterns emerge with notable cross-national consistency. Older males show robust effects across countries. Male 15-18 exhibits the largest effect sizes in both Austria (d = 0.77) and Germany (d = 0.90) medium-large to large effects by conventional standards. High-maturity older males predict approximately 3–3.5 months older than low-maturity peers within the same age band. This convergent finding across two independent samples provides strong evidence that digital maturity configurations meaningfully differentiate developmental status among older male adolescents. The German effect size of 0.90 approaches one standard deviation of separation between terciles. Younger females show null effects across countries. Female 11-14 exhibits negligible effects in both Austria (d = -0.16) and Germany (d = 0.02). In Austria, the direction is actually reversed, high-maturity younger females predict slightly younger than low-maturity peers, though the magnitude is trivial. This convergent null finding is theoretically notable as the composite maturity index apparently does not translate into predicted age differentiation for this demographic, suggesting either that relevant developmental processes for younger females operate outside the DIMI measurement framework or that maturity configurations function differently across gender and developmental stage.

Older females show modest effects. Female 15-18 exhibits small effects in both countries (Austria d = 0.22; Germany d = 0.14), weaker than older males but directionally consistent. This pattern aligns with Layer 2 findings that older female development centers on emotional regulation, a single dimension, rather than broad maturity profiles. The composite index may underweight emotional regulation's contribution relative to its actual developmental salience for this subgroup. Younger males show divergent cross-national patterns. Male 11-14 exhibits a small-medium effect in Germany (d = 0.38) but a negligible effect in Austria (d = 0.10). This divergence represents the clearest cross-national difference and aligns with the differing Layer 2 confirmation patterns between countries.

Step 3: Decomposing Individual Developmental Profiles

The tercile comparisons in Step 2 summarize average separation, but averages can obscure substantial individual-level heterogeneity. As documented in Step 1, within-age prediction variance spans approximately 1.0–1.5 years even among adolescents of identical chronological age. This implies that some individuals deviate substantially from their chronological peers in predicted developmental age. Step 3 therefore moves inside the model. We select illustrative cases where predicted developmental age is markedly higher or lower than chronological age and decompose their predictions using SHAP (SHapley Additive exPlanations) waterfall analysis. The objective is not to generalize from individual cases, but to demonstrate what a large deviation looks like in the model's internal accounting (which dimensions contribute positive or negative increments and how they cumulate to produce the final prediction). Figures 2.15 and 2.16 present SHAP waterfall decompositions for selected extreme cases, illustrating how dimensional contributions cumulate to produce the final predicted age.

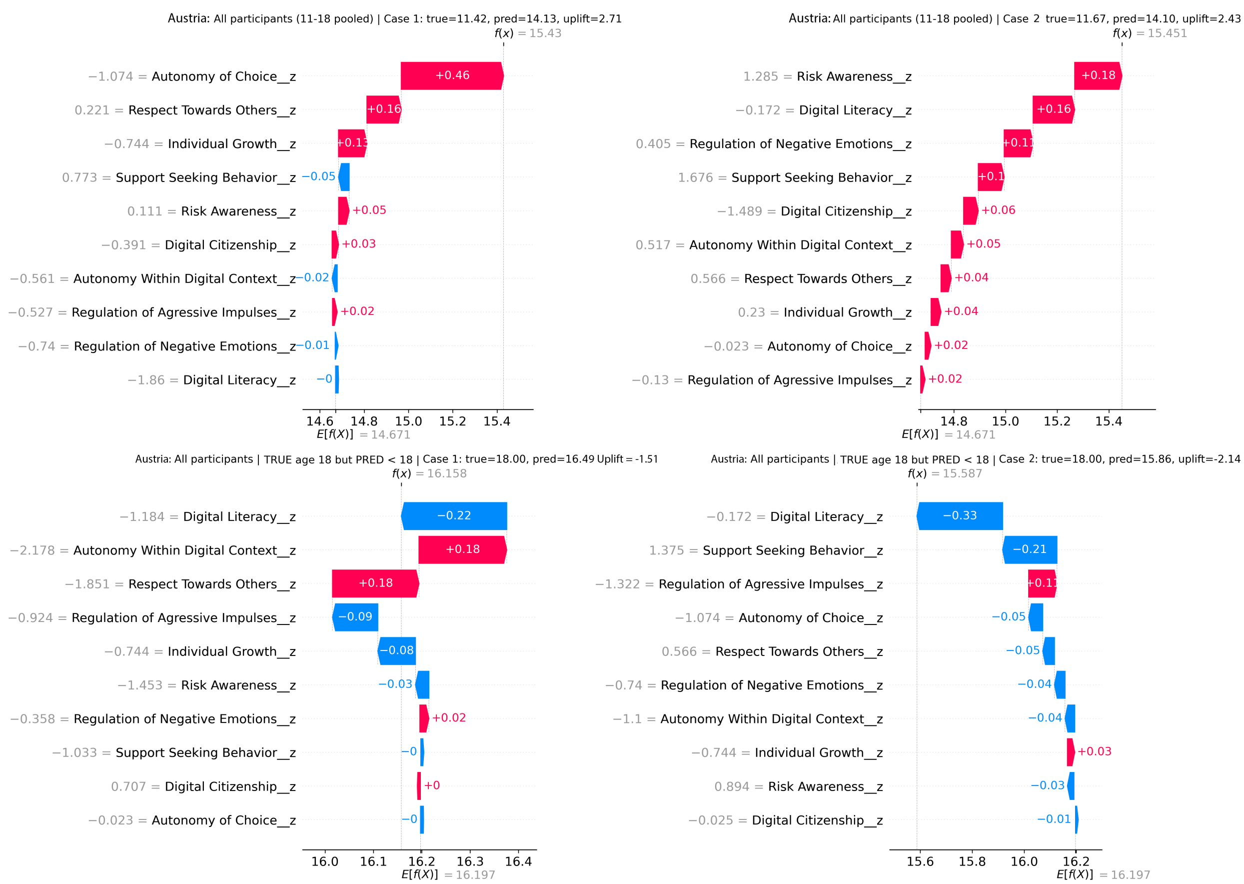

Figure 2.15: Austria Advanced and Delayed Cases SHAP Waterfall Plot

Austrian Cases: The Austrian advanced cases reveal two distinct pathways to developmental advancement. Advanced Case 1:an 11-year-old predicting 14.13 years (+2.71 year deviation),shows Autonomy of Choice contributing the largest positive effect (+0.46 years), with Respect Towards Others and Individual Growth adding smaller increments. Notably, this adolescent scores below average on Autonomy of Choice (z = -1.07), yet receives a pos-

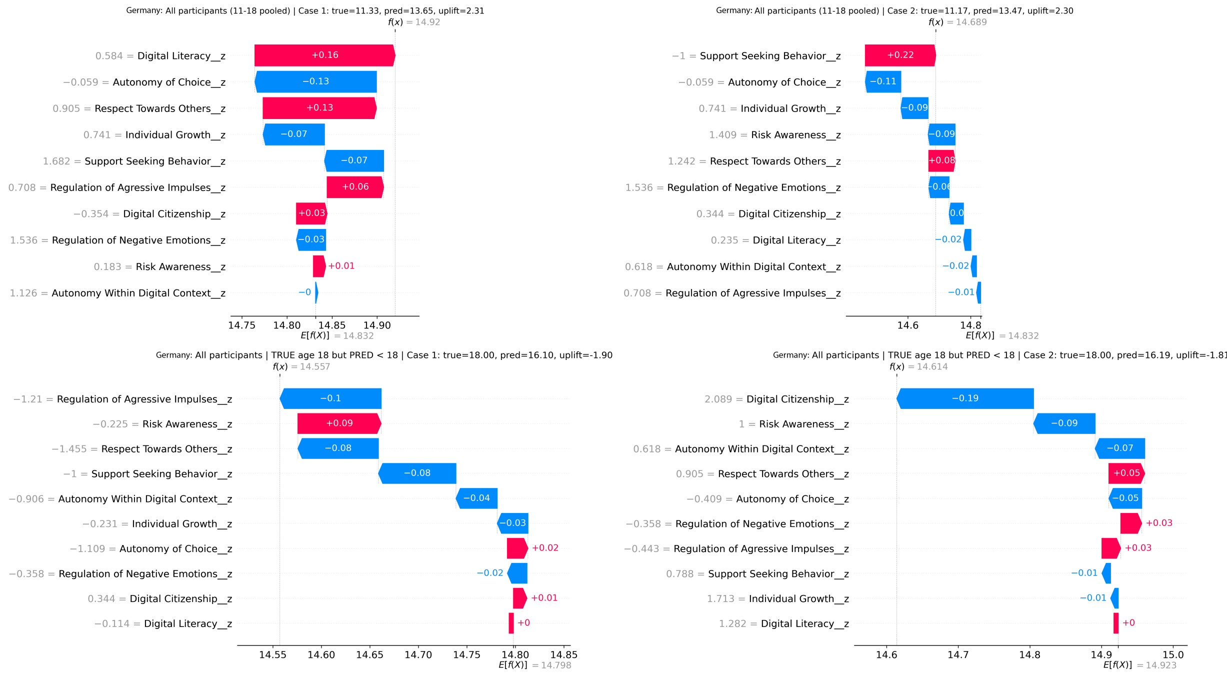

Figure 2.16: Germany Advanced and Delayed Cases SHAP Waterfall Plot

itive contribution; this counterintuitive pattern reflects the non-linear ALE functional form, where certain dimensions show elevated predicted age at both extremes. Advanced Case 2 shows a different configuration: distributed moderate strengths across Risk Awareness, Digital Literacy, Regulation of Negative Emotions, and Support Seeking Behavior each contributing +0.10 to +0.18 years. The Austrian delayed cases share a common driver: Digital Literacy deficits. Delayed Case 1 shows Digital Literacy contributing -0.22 years; Delayed Case 2 shows Digital Literacy contributing -0.33 years alongside Support Seeking Behavior (-0.21 years). In both cases, negative contributions from literacy and support-seeking dimensions anchor a younger developmental profile despite some positive contributions from other dimensions.

German Cases: German waterfall decompositions reveal both parallels and divergences from Austria. German Advanced Case 1 shows Digital Literacy (+0.16 years) and Respect Towards Others (+0.13 years) as primary drivers, similar to Austrian patterns. However, German Advanced Case 2 reveals a distinctive pathway: Support Seeking Behavior contributes +0.22 years while Digital Literacy contributes slightly negative (-0.02 years). This demonstrates that in Germany, prosocial dimensions can drive developmental advancement independently of literacy. The German delayed cases reveal the most striking cross-national divergence. Unlike Austria, where Digital Literacy deficits consistently dominated, German delayed cases show heterogeneous drivers. Delayed Case 1 shows Support Seeking Behavior (-0.19 years) as the largest negative contributor, with Digital Literacy contributing only -0.04 years. Delayed Case 2 shows Autonomy Within Digital Context (-0.13 years) and Digital Citizenship (-0.07 years) as primary negative drivers, while Digital Literacy actually contributes positively (+0.05 years).

This three-step analysis yields three key findings First, substantial within-age heterogeneity exists in both countries. Predicted age distributions span approximately 1.0–1.5 years within single chronological ages, and extreme cases show deviations of 2–3 years from actual age. This establishes that adolescents of identical chronological age occupy meaningfully different developmental positions as captured by digital maturity dimensions. Second, average maturity-based differentiation is modest but varies systematically by demographic. Mean tercile differences range from negligible (Female 11-14: essentially zero) to approximately 3.5 months (Male 15-18). Effect sizes exhibit remarkable cross-national consistency: Male 15-18 shows medium-to-large effects in both Austria (d = 0.77) and Germany (d = 0.90); Female 11-14 shows negligible effects in both countries (Austria d = -0.16; Germany d = 0.02). Third, these findings present a nuanced picture for age-based digital governance frameworks. Average tercile differences are modest, approximately 3.5 months even in the strongest subgroup, suggesting that for the typical adolescent, chronological age provides a reasonable approximation of developmental status. However, individual-level variance is substantial, and extreme cases show deviations of 2–3 years. Age-based thresholds assume developmental homogeneity within age cohorts, an assumption that holds reasonably well on average but fails for adolescents in the distributional tails.

2.4.4 Supplementary Analysis: Weekly Usage and Digital Maturity

To assess whether the maturity captured by the Random Forest model is reducible to usage volume, we conducted supplementary analysis comparing Week_Use distributions across maturity terciles within each subgroup (Country × Gender × AgeBand). Using the same out-of-fold maturity estimates as the main analysis, we classified adolescents into Low, Mid, and High maturity terciles and examined whether weekly usage differed systematically across these groups. This analysis addresses the possibility that 'maturity' as captured by the model simply reflects time spent online rather than qualitative developmental differences.

Differences in Week_Use between maturity terciles were modest and did not display a consistent pattern that would suggest maturity is reducible to usage intensity. While some subgroups showed statistically significant differences in weekly usage across terciles, the direction was not uniform as in several Austrian subgroups, higher-maturity adolescents reported slightly lower weekly usage, whereas German subgroups showed more heterogeneous patterns. Interestingly, correlations between Week_Use and the age acceleration (Predicted Age minus chronological age) were uniformly near zero across all eight subgroups (Spearman ρ ranging from 0.10 to +0.06, all $p > 0.07$), indicating that weekly usage does not predict the developmental signal captured by the maturity model (see Appendix for full results). These results suggest that the maturity captured by the DIMI-based model reflects qualitative developmental patterns rather than usage intensity. Given the cross-sectional design and the absence of a clear mechanistic relationship, we treat weekly usage as a descriptive correlate rather than evidence of a behavioral mechanism.

2.5 Discussion

Our findings reveal that digital maturity during adolescence unfolds through distinct developmental phases characterized by shifting dimensional salience, gender-differentiated pathways, and substantial within-age heterogeneity. Digital Literacy emerges as a robust developmental marker across early adolescence, confirming in five of eight subgroups with the strongest chi-square values in the study. Emotional regulation emerges as a uniquely female-specific frontier in late adolescence, replicating across Austria and Germany despite institutional differences. Within-age heterogeneity in developmental status is substantial at the individual level (predicted age deviations of 2–3 years) yet modest on average (approximately 3.5 months between maturity terciles), presenting a nuanced picture for age-based digital governance frameworks. We discuss theoretical contributions to IS research on digital capabilities and inequality, address methodological implications, derive actionable insights for practice and policy, and acknowledge limitations.

2.5.1 Theoretical Contributions

Re-conceptualizing Digital Capabilities as Developmentally Dynamic

Our findings fundamentally challenge how IS research conceptualizes ‘digital skills’, ‘digital literacy’, and ‘digital capabilities’, constructs central to understanding digital inequality, technology adoption, and value creation from IT use (Burton-Jones and Straub Jr, 2006; Marcolin et al., 2000). Where the dominant paradigm treats digital competencies as relatively stable individual differences or skills accumulating incrementally (Hargittai, 2001; Van Dijk, 2020), our results demonstrate this conceptualization is inadequate for adolescent populations, where underlying cognitive, social, and emotional capacities are themselves rapidly developing.

We introduce the concept of *phase-structured digital capability development*, where different dimensions become rate-limiting at different developmental stages. Digital Literacy dominates early adolescence (ages 11–14), exhibiting near-linear associations with developmental stage and confirming robustly across contexts with functional forms showing persistent positive derivatives (approximately 0.15–0.25 predicted years per standard deviation). However, literacy effects weaken by late adolescence as populations converge toward baseline competence German males 15–18 show no confirmation ($\chi^2 = 3.31$, $p_{FDR} = 0.791$)) suggesting literacy operates as a necessary but time-limited foundational capability. This ceiling effect aligns with developmental psychology research on skill acquisition trajectories (Gnambs, 2021; Fraillon et al., 2020). Supplementary analysis further supports this interpretation as weekly usage intensity does not display a stable relationship with the maturity signal in our sample, suggesting that maturity reflects qualitative developmental patterns rather than time spent online.

The phase-structured pattern resonates with developmental neuroscience research on adolescent brain maturation. Casey et al. (2008) dual systems model demonstrates that prefrontal regulatory systems mature more slowly than subcortical reward and emotion systems, creating a developmental window where emotional reactivity may outpace regulatory capacity. Crone and Dahl (2012) extend this by characterizing adolescence as a period of heightened social-affective engagement where goal-directed behavior becomes increasingly flexible but also susceptible to social and emotional influences. Our findings translate these neurobiological insights into the digital domain: Digital Literacy represents the cognitive-technical foundation maturing relatively early, while emotional regulation represents the later-developing capacity that becomes rate-limiting precisely when social media engagement intensifies.

This phase structure significantly extends existing IS frameworks on digital inequality. Van Dijk (2005) sequential stages of digital inequality propose a hierarchy but assume relatively stable agents progressing through stages. Hargittai (2001) foundational work demonstrated that access alone is insufficient, differences in skills and usage patterns create new forms of inequality. Our findings suggest that for adolescents, the agent is fundamentally changing: different capabilities become necessary as cognitive, emotional, and social systems mature. The 'skill' enabling beneficial use at age 12 (digital literacy) differs categorically from the capability required at age 17 (for females: emotional regulation). This developmental dynamism challenges static conceptualizations of digital competence pervading IS research.

The phase structure also extends debates about the 'third-level digital divide' focusing on outcomes from technology use (Van Deursen and Helsper, 2015). Our findings suggest a *developmental divide*: even with equivalent access and training, adolescents at different developmental stages face different binding constraints on effective use. This aligns with Livingstone and Helsper (2010) observation that online skills and self-efficacy interact to shape how adolescents balance opportunities and risks, but extends their framework by demonstrating that the relevant skills themselves shift across development. The Layer 2 confirmatory results show this empirically: Austrian females aged 11–14 show Support Seeking Behavior as developmentally salient ($\chi^2 = 12.51$, $p_{FDR} = 0.019$), while Austrian females aged 15–18 show Regulation of Negative Emotions ($\chi^2 = 18.31$, $p_{FDR} = 0.004$) as the defining developmental marker. Interventions designed to reduce digital inequality must therefore be developmentally calibrated, addressing the binding constraint at each phase rather than providing uniform literacy training.

Gender, Sociomateriality, and Differential Digital Vulnerabilities

Our most theoretically significant finding concerns the unique importance of emotional regulation for older female adolescents. In Austria, Regulation of Negative Emotions confirmed strongly (($\chi^2 = 18.31$, $p_{FDR} = 0.004$); in Germany, it achieved nominal significance ($p = 0.023$, $p_{FDR} = 0.094$)). Critically, no male subgroup in either country showed emotional regulation among top-ranked dimensions or approaching significance. This gender-specificity, replicated across two independent national samples with over 3,000 adolescents, represents robust evidence for gender-differentiated digital developmental pathways.

This finding requires theoretical explanation grounded in sociomateriality (Leonardi, 2012; Orlikowski, 2007), the mutual constitution of technology and social structures. Digital platforms afford different action possibilities and pose different challenges depending on users' socialized orientations, developed competencies, and social contexts shaped by gender (Faraj and Azad, 2012; Treem and Leonardi, 2013). Visual social media platforms, where adolescent females disproportionately engage (Anderson et al., 2018), afford continuous opportunities for upward social comparison (appearance, lifestyle, popularity), robustly linked to negative affect and psychological distress particularly for females (Twenge and Martin, 2020; Fardouly et al., 2015). Recent research on social media's psychological mechanisms strengthens this interpretation. Nesi and Prinstein (2015) demonstrate that social media transforms traditional peer processes by enabling constant social comparison and quantified feedback-seeking, with effects moderated by gender and social status. Their finding that these processes disproportionately affect girls' depressive symptoms aligns precisely with our observation that emotional regulation emerges as a binding constraint specifically for older females. Kross et al. (2013) provide experimental evidence that passive Facebook consumption predicts declines in subjective well-being, while Verduyn et al. (2017) systematic review identifies passive use and social comparison as the key mechanisms. Our findings extend this literature by demonstrating that not all adolescents are equally vulnerable, emotional regulation capacity moderates the translation of platform affordances into developmental outcomes.

From a sociomaterial perspective, platform affordances interact with gendered socialization patterns to create differential vulnerability. Girls socialized toward communion, relational orientation, and emotional expressivity (Rose and Rudolph, 2006; Chaplin and Aldao, 2013) encounter platforms architected to amplify social comparison, quantify peer feedback through visible metrics, and make relational dynamics hypervisible (Burrow and Rainone, 2017). This creates what we term an *affordance-vulnerability mismatch*: platform affordances systematically trigger emotional responses requiring sophisticated regulation to manage adaptively.

The developmental timing of this mismatch is critical. Blakemore and Mills (2014) characterize adolescence as a sensitive period for sociocultural processing, during which the brain shows heightened responsivity to social stimuli including peer evaluation, acceptance, and rejection. For females, whose social media engagement centers on these relational dynamics, the combination of heightened social sensitivity and still-developing regulatory capacity creates a window of particular vulnerability. The ALE curves reveal this pattern: for Austrian older females, Regulation of Negative Emotions shows a step-down functional form (high positive values at low z-scores dropping at mid-range), indicating that low emotional regulation is particularly diagnostic of younger developmental status. For German older females, the U-shaped pattern suggests different developmental dynamics but convergent directional effects. This finding extends how IS researchers conceptualize affordances. The dominant framing treats affordances as action possibilities that users may or may not actualize based on goals and capabilities (Volkoff and Strong, 2013; Majchrzak and Markus, 2012). Our findings suggest a complementary framing: affordances as *regulatory demands* that interact with users' developmental stage and socialized vulnerabilities. For adolescent populations, the question is not only "what can users do with this feature?" but "what does this feature demand of users' still-developing regulatory systems?" This reframing suggests that affordance theory, as applied to youth populations, requires explicit attention to the match between platform demands and user capacities, what we term *affordance-vulnerability alignment*.

Within-Age Heterogeneity: Challenging Assumptions of Developmental Homogeneity

Arguably our most consequential finding for IS theory and practice is the documentation of within-age heterogeneity. Predicted age distributions span approximately 1.0–1.5 years within single chronological ages, and extreme cases show deviations of 2–3 years from actual age in both countries. However, our findings also temper overly strong claims about maturity-based differentiation: average tercile differences are modest (approximately 3.5 months even in the strongest subgroup (Male 15–18: Austria $d = 0.77$; Germany $d = 0.90$))and effect sizes range from negligible to medium-large depending on demographic context.

This pattern of substantial individual-level variance combined with modest average effects resonates with emerging theoretical frameworks in media effects research. Valkenburg and Peter (2013) Differential Susceptibility to Media Effects Model (DSMM) proposes that media effects are conditional on individual difference variables moderating how users respond to media stimuli. Our findings provide empirical specification of this framework for adolescent populations: digital maturity dimensions represent the dispositional susceptibility variables determining whether platform affordances translate into developmental advancement or delay. Beyens et al. (2020) demonstrate this heterogeneity empirically, showing that social media effects on well-being vary substantially from adolescent to adolescent. Our within-age heterogeneity findings reveal the developmental substrate of this variation: adolescents differ not only in how they respond to platforms but in the regulatory capacities they bring to digital engagement.

This pattern presents a finer picture as for the typical adolescent, chronological age provides a reasonable approximation of developmental status, but age-based frameworks fail for adolescents in the distributional tails. This aligns with Belsky and Pluess (2009) differential susceptibility theory, positing that individual differences in environmental sensitivity create heterogeneous responses to identical environmental inputs. In our context, adolescents with regulatory deficits may be disproportionately affected by platform affordances overwhelming their coping capacities. Orben and Przybylski (2019) large-scale analysis finding small average associations between digital technology use and well-being is consistent with our modest average tercile differences, but as they note, small averages can mask substantial heterogeneity at distributional extremes.

The SHAP waterfall decomposition of extreme cases demonstrates what generates large developmental deviations. Developmental advancement can occur through multiple pathways (either concentrated dimensional strength or distributed moderate strengths), while delay often stems from a smaller set of dominant negative contributors differing across national contexts. In Austria, Digital Literacy deficits consistently anchor developmental delay; in Germany, delay stems from heterogeneous sources including prosocial and autonomy dimensions.

This challenges a fundamental assumption underlying most IS research, policy, and design involving youth populations: that chronological age adequately proxies for relevant capabilities, needs, and vulnerabilities. IS research on technology adoption has extensively examined age as a moderator (Venkatesh et al., 2003; Morris and Venkatesh, 2000), but treats age as a proxy for stable cohort membership or accumulated experience, not developmental stage. Our findings suggest that for adolescents, chronological age is a crude and often misleading proxy, within any single age year, we observe developmental spread equivalent to 2–3 years. This has important implications for person-technology fit frameworks (Goodhue and Thompson, 1995), suggesting that for adolescents, relevant 'user characteristics' are dynamically developing capabilities varying enormously within age cohorts.

Unexpected Findings and Boundary Conditions

While our theoretical framework anticipated phase-structured development and gender differentiation, several findings diverged from expectations and warrant theoretical attention. Most notably, German older males (15–18) showed complete null results across all four top-ranked dimensions, with chi-square values uniformly below 3.5 and p_{FDR} values exceeding 0.79. This comprehensive non-confirmation contrasts sharply with Austrian older males, where Digital Literacy confirmed strongly ($\chi^2 = 35.97$, p_{FDR} ¡ 0.001). Several interpretations merit consideration. First, German older males may exhibit unusually low developmental heterogeneity on DIMI dimensions, suggesting greater uniformity in digital maturity trajectories. Second, the DIMI instrument may not capture the most developmentally salient dimensions for this demographic as German adolescents may develop along dimensions not well-represented in the current measurement framework. Third, institutional differences between Austrian and German educational systems, including Germany's earlier tracking (age 10 versus Austria's age 14) and stronger media literacy curricula, may produce more uniform developmental trajectories by late adolescence. Hasebrink et al. (2009) cross-European analysis demonstrates substantial national variation in adolescent digital experiences, suggesting that the binding constraints on digital development may themselves be culturally and institutionally contingent. The reversed tercile ordering for younger females in both countries (high-maturity groups predicting at or below low-maturity groups) similarly challenges simple developmental narratives. This pattern suggests that the dimensions most predictive of older developmental status among younger females may not be the same dimensions that constitute the overall maturity index. For younger females, developmental advancement may occur through pathways perhaps involving social sophistication or relational competencies that the DIMI does not fully capture or that operate differently from other dimensions when aggregated.

These boundary conditions have important implications for theory development. Turel and Serenko (2012) demonstrate that the same technology features can produce both benefits and harms depending on user characteristics and usage patterns. Our findings extend this insight developmentally as the same digital maturity dimensions may operate as developmental markers in some demographic-cultural contexts but not others. This suggests that theories of digital capability development must be specified with explicit boundary conditions regarding gender, developmental stage, and institutional context rather than assumed to generalize universally. Moreover, These boundary conditions suggest two refinements to our theoretical framework. First, phase-structured development may be more pronounced in contexts where digital maturity trajectories are less institutionally scaffolded; Germany's stronger media literacy curricula may compress developmental variation by late adolescence, leaving less heterogeneity for DIMI dimensions to capture. Second, the composite maturity index may obscure dimension-specific developmental pathways, particularly for younger females whose advancement may occur through relational or social dimensions not well-represented in the current operationalization. Future research should examine whether alternative dimensional weightings or gender-specific maturity indices better capture developmental differentiation across demographic subgroups.

2.5.2 Practical Implications

Platform Design: From Engagement Optimization to Developmental Sensitivity

Our findings create ethical imperatives and practical guidance for platform designers. If emotional regulation is the binding constraint for older female adolescents, platforms could incorporate features supporting rather than overwhelming regulation. The ALE curves reveal that low emotional regulation is particularly diagnostic of younger developmental status among older females, suggesting platform design should attend to users with regulation deficits rather than treating all users equivalently.

We propose four design principles grounded in our findings. First, implement friction mechanisms slowing reactive behaviors: mandatory cooling-off periods before posting emotionally charged content, prompts encouraging reconsideration, or reminders after extended comparison-inducing scrolling (Lyngs et al., 2019). Lukoff et al. (2018) demonstrate that users themselves often desire such friction mechanisms to support self-regulation, suggesting that developmentally-sensitive design need not conflict with user preferences. Second, provide user control over emotion-triggering affordances: user-controlled feeds prioritizing close friends, limiting comparison opportunities, or filtering appearance-focused content rather than algorithmic feeds optimized solely for engagement. Third, incorporate developmental adaptation: assess users' digital maturity profile and adapt affordances accordingly, with less mature users receiving enhanced support mechanisms. Fourth, enable transparency about design choices: communicate how platform features may interact with regulatory capacities. Recent regulatory developments like the EU Digital Services Act requiring platform risk assessments for minors (European Commission, 2025) and the UK Online Safety Act mandating age-appropriate design increasingly mandate such responsibility.

Implications for Digital Education

Current digital education largely treats competence as fixed curriculum delivered uniformly to age groups. Our findings raise questions about this approach, though we acknowledge the practical constraints educators face. The confirmation of Digital Literacy across early adolescent subgroups, combined with the emergence of emotional regulation as salient for older females, suggests that curricula emphasizing only technical skills may be incomplete, particularly for older adolescents navigating emotionally challenging digital environments.

The within-age heterogeneity we identify (2–3 year developmental spread within single age groups) points toward potential limitations of age-grade approaches, though we recognize that maturity-based differentiation introduces its own challenges around assessment validity, resource allocation, and potential stigmatization. Przybylski and Weinstein (2017) finding that moderate digital engagement associates with optimal well-being (the "Goldilocks hypothesis") suggests that educational interventions should focus not on minimizing technology use but on developing the regulatory capacities that enable beneficial engagement.

Our findings specify which capacities matter at which developmental stages: Digital Literacy foundations in early adolescence, emotional and behavioral regulation in later adolescence, with gender-differentiated emphasis.

Our findings do not prescribe specific curricular solutions but rather suggest that educators and curriculum designers might benefit from considering: (i) whether phase-appropriate content could complement uniform approaches; (ii) how social-emotional competencies might be integrated with technical skill development; and (iii) whether diagnostic assessment could identify students whose developmental profiles diverge substantially from age-based expectations.

Considerations for Policy and Regulation

Most digital policy uses chronological age as the primary regulatory instrument like data collection prohibitions under age 13 (COPPA), platform age minimums, age-stratified educational mandates. These frameworks provide administrative simplicity and legal clarity, virtues that should not be dismissed. Our findings speak to both the reasonableness and the limitations of such approaches.

On one hand, the modest average tercile differences we observe (approximately 3.5 months between high- and low-maturity groups) suggest that chronological age is not fundamentally flawed as a population-level proxy for developmental status. For typical adolescents, age-based thresholds may function adequately. On the other hand, the substantial individual-level variance (2–3 year deviations in predicted developmental age) indicates that categorical age thresholds will systematically misclassify adolescents in the distributional tails, both advanced adolescents who may be unnecessarily restricted and delayed adolescents who may be insufficiently protected.

Our results do suggest that policymakers might consider how existing frameworks could accommodate developmental variation without abandoning the practical benefits of age-based defaults. Recent regulatory developments, such as the EU Digital Services Act's requirement that platforms ensure "a high level of privacy, safety and security of minors" (European Union, 2022a) and the emerging emphasis on age-appropriate design appear to move toward recognizing that age alone may be insufficient, though implementation challenges remain substantial.

2.5.3 Limitations and Future Research

Our findings should be interpreted in light of several limitations, some of which our research design partially addresses while others point toward productive future research directions. First is the cross-sectional design. Our most significant limitation is the cross-sectional nature of both samples. While we observe age-related patterns consistent with developmental theory, we cannot establish temporal precedence or rule out cohort effects. The phase-structured patterns we identify are consistent with developmental change but could alternatively reflect stable cohort differences or selection effects. We partially address this concern through cross-national replication across Austria and Germany strengthening confidence that observed patterns reflect genuine developmental processes rather than sample-specific artifacts. Nevertheless, longitudinal data remain essential for validation. Future research should employ accelerated longitudinal designs enabling within-person developmental trajectories, tests of whether early literacy predicts later prosocial and emotional regulation development, and stronger causal inference about developmental sequencing.

Second is the self-report measurement. The DIMI instrument relies on self-report assessment of digital maturity dimensions, which is subject to social desirability bias and requires meta-cognitive awareness that itself develops across adolescence. We mitigate this concern by using a validated instrument with established psychometric properties (Laaber et al., 2023), and the convergent patterns across independent national samples suggest that measurement artifacts alone cannot explain our findings. However, convergent validation with behavioral measures would strengthen claims. Future research combining self-report with performance-based assessments (e.g., ICILS digital skills tests), behavioral indicators from digital trace data, and outcome validation (testing whether maturity profiles predict well-being, psychological distress, or problematic use) would enable triangulation across measurement approaches.

Third, is the geographic scope pf this research. Our samples are restricted to Austria and Germany, two Western European countries with similar educational systems, digital infrastructure, and cultural contexts. While limiting generalizability, this scope provides a conservative test as the cross-national divergences we document between these relatively similar countries (Austria confirming 10 of 16 combinations versus Germany's 3 of 16) suggest that even broader variation would emerge across more diverse contexts. Expansion to non-Western societies, contexts with different digital infrastructures, and regions with different platform ecosystems would test boundary conditions and reveal whether dimensional salience depends on dominant platforms and usage patterns in each context. Lastly, our analysis identifies associations but cannot establish causal mechanisms. The emotional regulation finding for older females, for instance, could reflect that (a) developing emotional regulation capacity enables more mature digital engagement, (b) digital experiences that challenge regulation impede developmental progress, or (c) both capacities are driven by common underlying factors. Experimental research, platform affordance manipulations, emotion regulation interventions, or educational program comparisons, would enable stronger causal claims about how capabilities develop and whether interventions can modify trajectories.

2.5.4 Conclusion

This research makes thee primary contributions to IS Literature. Theoretically, we introduce phase-structured digital capability development, demonstrating that effective IT use requires different capabilities at different developmental stages and challenging static conceptualizations of digital skills. We extend affordance theory by showing how platform features interact with users' developmental stages and gendered vulnerabilities to create differential capability demands. And we document within-age heterogeneity that presents a nuanced picture for age-based frameworks which is modest on average but substantial in the tails. Methodologically, we demonstrate how interpretable machine learning can serve explanatory goals when paired with confirmatory inference and appropriate interpretability tools. We also provide systematic cross-national documentation of how digital maturity dimensions relate to developmental stage across adolescence, revealing universal patterns (Digital Literacy's foundational role), gender-differentiated pathways (emotional regulation's female-specific salience), and important contextual variation. Practically, we derive actionable implications for platform design, digital education, and policy that account for developmental complexity.

Digital maturity during adolescence is not a simple progression from incompetence to competence but rather a phase-structured developmental process where different dimensions become rate-limiting at different stages and for different demographic groups. For IS research, these findings demonstrate the importance of considering developmental context when studying technology use and outcomes among youth populations. Technologies do not affect adolescents uniformly, rather, effects depend critically on which capabilities users possess, which capabilities platforms demand, and how well these align. [1]

[1]The research is funded by the European Union's Horizon 2020 research and innovation programe under grant agreement No 870578, available here: https://cordis.europa.eu/project/id/870578

Designing for Data Democratization: Mechanisms from the EU Data Act

Abstract

Data democratization seeks to broaden who can access, use, and benefit from data while maintaining safeguards for individuals and society. To achieve data democratization, specific policies need to be formulated taking into account all stakeholder needs. However, policies at global level are currently limited. We draw on the case of the European Union as an illustrative example of value-sensitive policy design for data democratization which has been based on stakeholder needs. We analyze 102 stakeholder recommendation letters and legislative text from the Data Act process. We explore convergences and tensions among stakeholders and the EU across and then identify four democratization meta-requirements: equitable access, protected openness, pragmatic interoperability, and participatory accountability. We show how the Act assembles policy mechanisms that serve these requirements and where tensions persist. We conclude with design lessons for other economies, contrasting the EU's balanced, rights-forward approach with the more market-led US debate and the more state-centric model often associated with China. Our contribution is a how-to roadmap for policy design toward data democratization, grounded in a concrete, large-scale case that connects values to implementable mechanisms.

Key Words: Data democratization; EU Data Act; Value Sensitive Design; Data governance; NLP; Policy mechanisms

3.1 Introduction

Data governance faces a fundamental paradox: organizations require expansive data access to drive innovation (Janssen et al., 2012), while societies demand safeguards against algorithmic harm (Mikalef et al., 2022). Today, across economies, the stakes of data governance have shifted from organizational efficiency to societal choice. Data democratization emphasizes the aspiration that data and its value should not be confined to a few actors, but broadly accessible under fair and accountable conditions (Lefebvre et al., 2021; Meng, 2024; Janssen et al., 2012). Yet, practical guidance on how to design policies that realize this aspiration remains thin. Firms emphasize access and flexibility to innovate; publics demand fairness, privacy, and accountability (Mikalef et al., 2022; European Commission, 2018). Small and medium enterprises are especially vulnerable to data concentration and vendor lock-in, and many economies now seek frameworks that expand access without eroding legitimate protections.

In this context, the EU provides a useful case for examining how data democratization aspirations are translated into concrete policy mechanisms. Building on the European Strategy for Data, the EU Data Act aims to increase fair access to and use of data and to curb structural lock-in, while safeguarding privacy, trade secrets, and non-discrimination (European Commission, 2018, 2021, 2022b). Though prior work examines organizational data ethics (Cecez-Kecmanovic, 2019) and policy implementation (Dutton, 2014), we currently lack empirical insights on how stakeholder values, such as innovation, fairness, and accountability, converge to or conflict with regulatory priorities during regional policymaking. This gap limits our understanding of how data democratization is operationalized practice, particularly in the context of balancing innovation with ethical imperatives (Wachter and Floridi, 2017). The EU's approach is informative beyond the European context, as it offers an integrative perspective into how diverse economic stakeholders, public interest and regulatory objective interact within a large and heterogeneous market (Williams, 2010). In this paper, our goal is to understand *'How can policymakers design for data democratization in complex data ecosystems, and which policy mechanisms operationalize its core requirements?'*

We address this question by examining how competing stakeholder values are translated into implementable policy instruments in the EU context. We adopt Value Sensitive Design (VSD) as an analytical lens for policy analysis, conceptualizing legislation as a governance artifact that considers value tradeoffs through rights, exceptions, standards, and contractual controls (Friedman et al., 2006, 2019). Methodologically, we employ a sequential mixed-methods design: First, a transformer-based NLP classifier maps paragraphs to VSD values across the full corpus of 102 stakeholder consultation letters and the legislative text. Then, a thematic analysis develops in-depth interpretations of value tensions and how they are addressed within the Act. This approach allows us for a more comprehensive understanding of the values associated with the stakeholders and the EU (Venkatesh et al., 2013; Mingers, 2001).

In this study, we first identify notable differences in value emphasis: stakeholders tend to focus on access-enabling concerns such as usability, ownership and property, and trust, while the legislative text places weight on safeguard-oriented concerns such as accountability, privacy, consent, and bias mitigation. Second, moving beyond individual values, we derive four meta-requirements that operationalize data democratization in policy: *equitable access* (entitlements that weaker actors can actually use), *protected openness* (data sharing under enforceable safeguards), *pragmatic interoperability* (portability and switching embedded in infrastructure arrangements), and *participatory accountability* (clear, auditable responsibility pathways for recourse). We show how these requirements are instantiated in the case of the EU and where they align with, or diverge from, stakeholder expectations.

Building on these results, the paper contributes to IS theory and practice in three ways. First, it develops a mechanism-based account of democratization that is observable in law and contracts, complementing organizational perspectives that emphasize access, literacy, and culture with policy-level design logics. Second, it extends VSD to techno-regulatory design by showing how value clusters become enforceable obligations, which clarifies the translation path from value identification to institutional embedding. Third, it offers an AI based method for value detection at scale that connects NLP outputs to qualitative explanation, supporting cumulative inquiry on policy design in digital ecosystems.

For practice, the EU case indicates that democratization advances when rights are coupled to capabilities for use, when openness is paired with guardrails, when interoperability is built as shared infrastructure, and when accountability is usable for the least powerful. We turn these insights into practical policy recommendations that can help step-by-step reforms in places that do not adopt a single horizontal, all-purpose law.

3.2 Literature Review

3.2.1 Data democratization and value–sensitive policy

Across the last decade, data has shifted from an internal asset to a societal infrastructure that organizes markets, public services, and everyday life. *Data democratization* captures the ambition that access to, and benefits from, data should be broadened across actors while respecting fundamental rights and legitimate interests. In IS, democratization is connected to reducing gatekeeping around data access and capability while maintaining prudent risk controls and literacy so that insight generation is not confined to technical elites (Lefebvre et al., 2021; Gregory et al., 2018). At a policy level, the same objective is complicated by heterogeneous actors, asymmetric power, and cross–border flows of data. Where organizations can re-allocate decision rights internally, regulators must translate normative aims such as fairness, accountability, and non–discrimination into enforceable mechanisms that operate across ecosystems.

Here, a mechanism view is useful: democratization at scale is not a single instrument but a bundle that includes rights of access and portability, protections against exploitation, interoperability and switching to lower structural frictions, and accountability that is legible and enforceable for weaker parties. These levers must cohere so that opening access does not erode privacy or trade secrets, and safeguarding rights does not recreate structural lock–in that excludes smaller actors. This policy design problem is value–laden by nature, which motivates a design logic that makes value trade–offs explicit and connects them to concrete rules.

3.2.2 Data democratization in IS and policy

IS research has long conceptualized data governance as the allocation of decision rights and accountability for data–related processes (Khatri and Brown, 2010) and has mapped its organizational mechanisms and outcomes (Abraham et al., 2019). Recent work identifies democratization enablers within firms as broad data access, self–service analytics, data literacy, collaborative culture, and recognition of data's value in decision making (Lefebvre et al., 2021). These enablers dismantle silos and allow non–specialists to work productively with data, provided that guardrails are in place. Yet at ecosystem level the levers change. Concentration of data assets and capability, unfair contracting power, and switching frictions can exclude SMEs even if data is nominally available. Open data alone is insufficient for industrial data that is co–generated by devices and services and often entangled with IP and confidentiality. Democratization at this level requires a rules–based environment that lowers structural barriers to participation without dissolving legitimate protections.

Public policy in the EU has embraced this ecosystem perspective. The *European Strategy for Data* set the objective of a single market for data with common data spaces that increase availability and re–use in trustworthy ways (European Commission, 2020). The *Data Governance Act* (DGA) focuses on trust and intermediaries for sharing and altruism, and the *EU Data Act* focuses on fair access and use across sectors (European Union, 2022b, 2023). Together, these instruments aim to rebalance access, reduce lock–in, and enable cross–sector data flows while sustaining fundamental rights. From a democratization standpoint, the bundle addresses four structural frictions: access asymmetries, contracting power imbalances, technical and commercial switching costs, and weak recourse for smaller actors. In parallel, the literature on data spaces and meta–organizations highlights that cross–firm collaboration around shared data requires both technical compatibility and institutional arrangements that reduce transaction costs and distribute benefits more evenly (Guggenberger et al., 2025). The policy and IS streams therefore converge on the same design challenge: democratization requires jointly designed technical and institutional mechanisms, not declarations of openness alone.

3.2.3 VSD as a design logic for policy

To translate democratization aims into implementable policy mechanisms, we adopt Value Sensitive Design as our analytical framework, which makes value trade-offs explicit and links them to concrete design choices. Value Sensitive Design (VSD) provides a principled way to connect human values to design choices through conceptual, empirical, and technical investigations (Friedman et al., 2013, 2019). Canonical VSD values include privacy, ownership and property, accountability, universal usability, autonomy, informed consent, trust, and freedom from bias (Friedman et al., 2006). Although VSD is often applied to artifacts and IS development, the same method extends naturally to policy treated as a designed artifact. Conceptually, VSD surfaces value priorities and tensions among stakeholders. Empirically, it examines how different communities understand and prioritize those values. Technically, it links value commitments to specific design moves.

In the context of data democratization, VSD supplies two benefits. First, it gives a stable vocabulary to reason about what ‘opening up’ should and should not entail. Universal usability, for example, pushes toward access and capability; privacy, freedom from bias, and informed consent push toward safeguards; ownership and property points to defensible interests in co–generated and processed data; accountability links the whole to traceable responsibilities and recourse. Second, VSD invites explicit trade–off analysis. Mechanisms that satisfy one value can degrade another if not co–designed. Democratization policy must therefore assemble mechanisms whose value consequences are understood in combination. We then apply this lens to the EU Data Act, a horizontal regulation whose mechanism bundle explicitly targets access, protection, interoperability, and accountability.

3.2.4 The EU Data Act as an Illustrative case

The EU is a salient setting for studying democratization–oriented data policy because its instruments are explicitly crafted to widen participation while sustaining rights and fair

Table 3.1: VSD values definition

Value	Definition
Human welfare	Refers to people's physical, material, and psychological well-being.
Ownership and property	Refers to a right to possess an object (or information), use it, manage it, derive income from it, and bequeath it.
Privacy	Refers to a claim, an entitlement, or a right of an individual to determine what information about himself or herself can be communicated to others.
Freedom from bias	Refers to systematic unfairness perpetrated on individuals or groups, including pre-existing social bias, technical bias, and emergent social bias.
Universal usability	Refers to making all people successful users of information technology.
Trust	Refers to expectations that exist between people who can experience good will, extend good will toward others, feel vulnerable, and experience betrayal.
Autonomy	Refers to people's ability to decide, plan, and act in ways that they believe will help them to achieve their goals.
Informed consent	Refers to garnering people's agreement, encompassing criteria of disclosure and comprehension (for "informed") and voluntariness, competence, and agreement (for "consent").
Accountability	Refers to the properties that ensure that the actions of a person, people, or institution may be traced uniquely to the person, people, or institution.
Identity	Refers to people's understanding of who they are over time, embracing both continuity and discontinuity over time.
Calmness	Refers to a peaceful and composed psychological state.
Environmental sustainability	Refers to sustaining ecosystems such that they meet the needs of the present.

Note. Definitions adapted from (Friedman et al., 2006).

competition. The *Data Act* entered into force in January 2024 and applies from September 2025, alongside phased obligations for some providers (European Union, 2023; Clifford Chance, 2023; Skadden, Arps, Slate, Meagher & Flom LLP, 2025). We treat the Act as democratization–in–practice because it assembles a mechanism bundle that targets known ecosystem frictions and ties those mechanisms to values recognizable in the VSD vocabulary.

First, the Act introduces user access and portability for data generated by the use of connected products and related services, enabling users to share such data with third parties on fair, reasonable, and non–discriminatory terms (European Union, 2023). This addresses access asymmetries and supports universal usability, while preserving trade secrets and privacy through protective conditions. Second, the Act constrains unfair terms in B2B data sharing where bargaining power is asymmetric, and it clarifies liability and contractual transparency. These provisions operationalize accountability in a way that is legible to weaker parties, especially SMEs. Third, the Act reduces structural switching costs by imposing switching and interoperability obligations on data processing services, which mitigates vendor lock–in and supports contestability in downstream services (Wilmer Hale, 2023). Fourth, the Act provides narrowly tailored business–to–government access in situations of exceptional need, subject to strict necessity, proportionality, purpose limitation, confidentiality, and compensation conditions (EU Data Act Chapter 5, 2024). Finally, where smart contracts are used to automate data sharing agreements, the Act sets essential requirements around robustness, access controls, safe termination, and auditability to prevent irreversible or opaque execution

(EU Data Act Chapter 36, 2024).

Taken together, these mechanisms speak directly to democratization's core, that is, equitable access without naive openness, protected sharing that does not erode rights or investment incentives, pragmatic interoperability that lowers switching costs across providers, and participatory accountability that gives weaker parties recourse in practice. The EU case is analytically valuable because its mechanism mix is explicit and horizontal, which makes it possible to examine how stakeholder value claims align with, or diverge from, the regulator's embedded value commitments.

In what follows, we use a VSD–guided, mixed–methods approach to compare value salience and interpretation in stakeholder submissions and the legislative text, and we interpret how the mechanism bundle resolves or leaves open key tensions in the road to data democratization.

3.2.5 Research Context & Dataset

We use a mixed–methods design: a quantitative NLP strand that detects value patterns in texts, followed by a qualitative strand that interprets those patterns in context. The roadmap was released in May 2021, with stakeholder feedback invited until June 2021. The European Commission then ran an Open Consultation from June to September 2021, inviting organizations and individuals from the EU and third countries to submit views and recommendations on topics including B2G and B2B data sharing, smart contracts, intellectual property, and cloud data sharing. This consultation informed the first draft of the Act, released on 22 February 2022 as a horizontal regulation on fair access to and use of data[1] (European Commission, 2022a). Our analysis focuses on the letters submitted during the open consultation that preceded this initial draft. In total, 449 stakeholders completed the initial survey and 102 submitted recommendation letters. Contributors included public bodies, NGOs, and private firms (e.g., UNICEF, Airbus, IBM, Salesforce, Vodafone Group, BBC, Telefónica). The analysis examines how stakeholder positions extend beyond intra–organizational interests toward ecosystem interdependencies and shared responsibilities in ethical data practices. The final corpus comprises 1,035 paragraphs and 279,165 words drawn from the stakeholder recommendations and the *EU Data Act*.

3.2.6 Mixed-Methods Design

We employ a sequential mixed-methods design that aligns with Value Sensitive Design's three-phase logic (Friedman et al., 2019). The first phase (quantitative NLP value mapping) precedes and guides the empirical phase (qualitative thematic analysis of stakeholder and legislative excerpts). This sequence addresses three challenges in our techno-regulatory setting: (i) bridging scale across 102 stakeholder letters and depth in a corpus of 279,165 words, (ii) connecting value frequency patterns with contextual meaning, and (iii) balancing stakeholder perspectives with regulatory intent. Our approach follows established IS work on mixed methods (Venkatesh et al., 2013; Mingers, 2001) while tailoring it to policy text analysis. The quantitative strand assigns each paragraph to one of 12 VSD values and computes

[1] After the first draft, the Commission accepted stakeholder comments between 14 March and 13 May 2022. The final Data Act entered into force on 11 January 2024 and will apply from September 2025.

value frequencies and co-occurrences across stakeholders and the Data Act; the qualitative strand then examines how these values are articulated in stakeholder narratives and embedded in legislative mechanisms (Vaast et al., 2017; Zhang, 2017; Polyviou et al., 2024). This approach provides greater completeness, providing a more comprehensive understanding of the phenomenon under research(Venkatesh et al., 2013).

Phase I: Quantitative Strand

We use Natural Language Processing to analyze complex policy and consultation texts and to extract semantic signals at scale (Li et al., 2021; Katz et al., 2023; Sleimi, 2018; Manning, 1999). For the *VSD Value Classifier*, we collected the recommendation letters and the EU Data Act and prepared them for analysis. The letters are available via the European Commission's website, and the Act is published on the EU's official portal. Our goal was a *VSD Value Classifier* that categorizes paragraphs into the 12 VSD values (Friedman et al., 2008, 2014, 2019). We started to construct a supervised training set from foundational VSD definitions and examples. By foundational VSD definitions we refer to the descriptions of each value and their illustrative examples as articulated in Friedman et al. (2008, 2014, 2019) work on VSD, including the formal value definitions and conceptual elaborations. We compiled these definitions to form a base corpus of the 12 VSD values. We then enriched this base corpus with additional variations (definitions, short scenarios, and synonyms) that reflect the type of phrasing found in legal and policy documents, so that the classifier would not overfit to textbook-style formulations. To increase robustness to wording differences, we applied RoBERTa-based data augmentation. We used a pre-trained RoBERTa language model to generate paraphrases and controlled lexical variations of the original value-labelled sentences (for instance by masking selected words and letting the model predict context-appropriate alternatives), thereby creating new training examples that preserve the underlying value label while exhibiting different surface forms. This augmentation helps the classifier recognise the same value when it is expressed in slightly different legal, technical, or organizational language. The final supervised training set contains 1,320 entries (49,848 words), comprising 110 variations per VSD value. Table 3.2 provides an overview of the complete algorithm used to construct and train the VSD Value Classifier.

Table 3.2: Algorithm overview for VSD value classifier and further analysis

Step	Description
1. Building the VSD Value Classifier	Constructing a base dataset for the VSD values from academic sources. Augmenting the base dataset using GPT3.5 for diverse VSD value representation. Enriching the dataset using large language model (RoBERTa).
1.a) Using Pre-Trained Model	Use of pre-trained model, EUBERT pretrained on a corpus from the European Publications Office for detailed understanding of textual details.
1.b) Model Training	Train the EUBERT model on the constructed dataset for VSD value classification.
1.c) Classification of Paragraphs	Classification of each paragraph from all the EU stakeholder recommendation letters and the EU Data Act.
2. Eliminating Low Probability	Exclude paragraphs classified with probability of belonging to any VSD below 50%.

We fine-tuned the pre-trained EUBERT model on this dataset to capture domain-specific regularities in EU legal and policy language (Campion, 2023). The split was 80–20, yielding approximately 39,878 training words and 9,965 test words. On the held-out set the classifier achieved 98.1% accuracy, F1 of 98%, and precision of 98.2%. Following established practice, paragraphs with predicted value probability below 50% were excluded from downstream analysis (Zou et al., 2016). Using the trained model, we labeled paragraphs from the 102 stakeholder letters and from the Data Act, and we computed value frequencies and co-occurrences for each source. Table 3.2 summarizes the pipeline; full details appear in Appendix A.3.2.

Phase II: VSD Guided Qualitative Analysis

While Phase I enabled us to identify the most salient VSD values arising from the stakeholders' letters and the Data Act, in Phase II we focus on complementing this picture with in-depth, qualitatively grounded insights into how these values are articulated in context. We concentrate on the top five stakeholder values, as these are found to be the most important values for the stakeholders in our analysis. Building on the results of the quantitative phase, we generated separate documents for each of these values. Each document consolidated all relevant stakeholder paragraphs classified into the respective value, and we followed a similar procedure to construct parallel documents that contained the corresponding EU Data Act text extracts.

The analysis process involved all authors assessing each value-specific document and reflecting on how the VSD value was expressed in the stakeholders' paragraphs and in the EU Data Act text. We followed a thematic analysis approach for each document, as it offers a flexible way of identifying and organising patterns within empirical material (Griva et al., 2023; Vaast et al., 2017; Braun and Clarke, 2006). In line with earlier work (e.g., (Griva et al., 2023)), our approach began with familiarisation with the data and verification that all stakeholder quotations and legislative passages carried their unique identifiers (for letters) or page references (for the Data Act). This VSD-guided thematic analysis allowed us to closely examine stakeholders' letters and the EU Data Act side by side for each value, focusing on quotations that addressed similar topics across the two document sets. For each value, we compared the themes identified from the stakeholders' letters with those identified from the Act, which enabled us to characterise how the same value is emphasised, elaborated, or qualified across different actors in the European data ecosystem. Read across values, these themes revealed recurring design problems that appear in multiple stakeholder groups and re-emerge in several provisions of the Act, pointing beyond isolated value tensions toward broader patterns in how data governance is being configured. In this way, Phase II deepens and contextualises the quantitative signals from Phase I by showing how salient values are articulated in practice and by preparing the ground for a higher-level synthesis in the discussion. The combined insights from Phase I and Phase II are presented in detail in the Findings section.

3.3 Findings

We examine how values around the *EU Data Act* are articulated by stakeholders and embedded in the legislative text. Phase I reports quantitative signals by identifying the most salient VSD values in stakeholder recommendation letters and in the Act. Phase II then provides close interpretation of those signals using representative excerpts from both the EU and the stakeholders. This sequencing lets us first establish where emphases converge or diverge, and then explain how those emphases appear in practice.

3.3.1 The Most Salient VSD Values

Our analysis identifies the top VSD values for stakeholders and for the EU, with three of the top five differing across the two sources. Figure 3.1 shows, in descending order, the normalized counts of values after classification. For stakeholders, *universal usability* ranks highest, followed by *accountability*, *trust*, *ownership and property*, and *environmental sustainability*. In the EU text, *accountability* ranks highest, *universal usability* is second, and *privacy*, *informed consent*, and *freedom from bias* are prioritized ahead of *trust*, *ownership and property*, and *environmental sustainability*. These distributions already indicates a notable difference in emphasis as the stakeholders appear to weight access and practical use alongside protection of legitimate interests, while the EU Data Act gives relatively greater weight to rights-forward safeguards in the context of sharing and use.

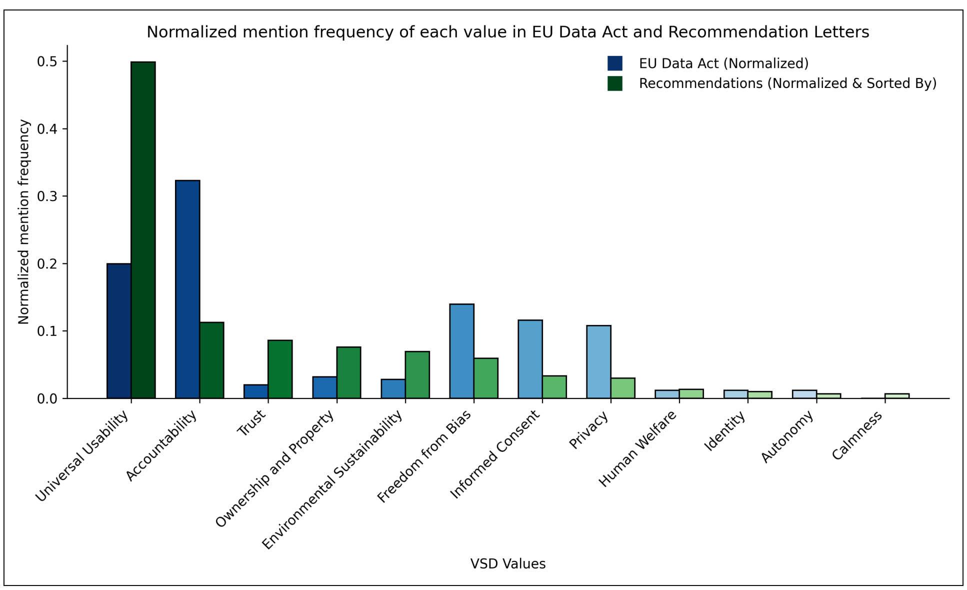

Figure 3.1: Normalized Stakeholders‘ Recommended Values vs EU Data Act Values

3.3.2 Qualitative Analysis Results: Exploring Stakeholder Values in the EU Data Act

Beyond identifying the most important values for stakeholders —*Universal Usability, Accountability, Trust, Ownership and Property,* and *Environmental Sustainability* (Figure 3.1), we further analyzed these values to uncover elements within stakeholder recommendations. We compared these insights with the corresponding values found in the EU Data Act to identify potential points of convergences and divergences. The qualitative analysis enables us to elaborate on how the stakeholders' and the EU Data Act interpret each value

Universal Usability

Stakeholders strongly emphasise *Universal Usability* as fair, collaborative data sharing across the ecosystem. They explicitly call for arrangements that protect the interests of all actors and avoid structural imbalances, arguing that the "interests of both data holders, e.g., large providers of IoT equipment, and [...] SMEs using IoT objects or providing supplemental services for them, are taken into account and that any imbalances favoring the former are actively avoided and rectified" [F2636044]. Universal usability is understood as a level playing field that "maximize[s] the collective benefits of data, both by sharing and by establishing fair user rights to data, as well as by developing skills to deliver the aspirations for the data economy" [F2668403]. Access is framed as a precondition for innovation in concrete sectors, for example that "access to vehicle-generated data is essential" in transportation [F2668337], and as a driver of Europe's digital position, since "the sharing and re-use of information assets not only strengthens Europe's position as a world leader in the digital economy, but also encourages simultaneously creating new opportunities for people and businesses" [F2669166]. Several letters present the Data Act as a reference point beyond Europe, suggesting that it could set "harmonized data sharing practices at an international level" [F2669494], and they underline that "the more balanced the business-to-business data sharing situations are, the more all parties benefit from innovation and productivity potential that a greater availability and use of data offers" [F2636044].

Stakeholders also fold data portability and interoperability into their understanding of universal usability. Portability is described as critical to balancing fundamental data protection rights with innovation along digital value chains [F2669274]. At the same time, contributors stress that "technical standards for data portability might be fundamentally different if we refer to health data or to electricity consumption" [F2669274], and they note that portability "might not always be practical" for certain content types such as images or videos [F2668966]. To keep requirements feasible, they argue for "industry-driven" technical specifications [F2669405] and warn that interoperability standards "imposed by large players [...] would have the consequences of preventing the smallest players from finding a place in the data market and from innovating" [F2669386]. In this way, universal usability is articulated as a combination of fair access, sector-sensitive standards, and governance arrangements that allow smaller actors to participate.

From the perspective of the EU Data Act, *Universal Usability* is articulated through concrete access, transparency and sharing obligations. The proposal requires that products be designed and manufactured, and related services provided, so that data generated by their use are “by default, easily, securely and, where relevant and appropriate, directly accessible to the user”, and it obliges manufacturers and service providers, before contract conclusion, to inform users in a “clear and comprehensible format” about the nature and volume of the data likely to be generated, whether they are generated continuously and in real time, how users may access them, whether the manufacturer or provider intends to use them or allow third-party use, and how the user may request that data be shared with a third party (Art. 3(1)–(2), pp. 40–41). Where users cannot directly access the data from the product, data holders must make available to the user the data generated by its use “without undue delay, free of charge and, where applicable, continuously and in real-time” on the basis of a simple electronic request (Art. 4(1), p. 41). Users also have a right, upon request, to have such data made available to a third party, free of charge to the user and “of the same quality as is available to the data holder”, and where applicable continuously and in real time (Art. 5(1), pp. 41–42). On the data holder side, the Act requires that data made available to recipients under legal or contractual obligations be provided on “fair, reasonable and non-discriminatory terms and in a transparent manner”, prohibits exclusive conditions except when requested by the user, and forbids discrimination between comparable categories of data recipients (Art. 8(1), (3)–(4), p. 44). Any compensation agreed for making data available must be reasonable and, where the data recipient is a micro, small or medium enterprise, “shall not exceed the costs directly related to making the data available” (Art. 9(1)–(2), p. 44).

Accountability

Stakeholders strongly advocate for *Accountability* in the Data Act, particularly in terms of clear, workable responsibilities for data handling. They emphasise that regulatory measures should harmonise procedures for data requests rather than impose new layers of obligation, given the complexity and speed of contemporary data flows. One submission states that “the Data Act should not lead to any new obligations for companies, but only harmonise the procedures for making and processing data requests across Member States, particularly when there is a public interest purpose” [F2669272]. Contributors highlight the practical difficulty of providing advance notice and extensive reporting in this environment, pointing to the “complexity of international data flows in a global economy, not to mention the speed at which certain access is required” and warning that additional reporting requirements could lead to “potential market disruptions” [F2669055]. They also call for clear legal pathways for “third-party data reuse and cross-border transfers”, advocating a risk-based approach that is sensitive to data content and business practicality, and opposing unnecessary restrictions on cross-border flows given their role in commerce and cybersecurity [F2669055; F2669298]. Within this framing, accountability is linked to clarity about who is responsible for what. Some letters favour a “transparent and shared responsibility between different parties, such as business users of digital technologies and providers of digital solutions and services” [F2669358], while others argue that “vendors should take responsibility for the data they manage on behalf of the customer [...] they should provide cloud technologies in such a way that customers can comply to regulations they are subject to” [F2669466]. Several stakeholders underline that “clear measures tailored to the specific uses are vital in order to remove legal uncertainty” [F2669226].

Accountability is also associated with enabling responsible data sharing through voluntary, contract-based arrangements rather than mandating disclosure in all cases. Stakeholders propose "facilitating the sharing of data based on free decision about voluntary data sharing or data sharing on a contractual basis" [F2669226] and see "voluntary model contract templates designed by the industry" as "an interesting tool to facilitate data sharing" [F2669272]. They suggest that voluntary sharing should be encouraged through incentives, especially in business-to-government settings. For example, B2G data-sharing schemes could include "reduction of transaction costs and perceived risks for data suppliers, tax incentives fostering standardisation of contractual provisions and interoperability of data formats" [F2669274]. Where data provision entails non-trivial effort, they recommend price mechanisms that recognise "costs incurred by companies to make data sets available, including time and resources spent preparing the data, adapting it to the specific request, and building the infrastructure necessary for the transmission of such data" [F2669480]. Across these contributions, accountability is understood as making roles, processes, and burdens explicit and proportionate, while preserving flexibility and voluntariness.

Accountability within the EU Data Act is implemented through a set of provisions that specify roles, safeguards and conditions for lawful data access and sharing. In its definitions, the proposal introduces *"smart contract"* as a "computer program stored in an electronic ledger system wherein the outcome of the execution of the program is recorded on the electronic ledger" (Art. 2(16), p. 40), anticipating programmable enforcement of sharing conditions. It then obliges data holders who are legally or contractually required to make data available to do so on "fair, reasonable and non-discriminatory terms and in a transparent manner", and it requires that such terms be contractually agreed and not exclude or vary the effect of users' rights, with unfair contract terms rendered non-binding (Art. 8(1)–(2), p. 44). To prevent abuse of negotiating power, the Act limits the possibility of exclusive access and prohibits discrimination between comparable categories of data recipients, shifting the burden of proof to the data holder where discrimination is alleged (Art. 8(3)–(4), p. 44). It further provides that, for micro, small or medium-sized enterprises, any compensation for making data available "shall not exceed the costs directly related" to doing so, and requires data holders to disclose the basis for calculating compensation so that recipients can verify compliance (Art. 9(2), (4), pp. 44–45).

The proposal also links accountability to traceability and user agency. It grants users the right to have data generated by the use of a product or related service made available, upon request, to a third party "without undue delay" and at the same quality as held by the data holder (Art. 5(1), pp. 41–42) and stipulates that third parties must process such data only for the agreed purposes, must delete them when no longer necessary, and must not use them to develop competing products or to profile individuals beyond what is necessary for the requested service (Art. 6(1)–(2), p. 43). To resolve disputes over terms and transparency in making data available, the Act requires Member States to certify dispute-settlement bodies that are impartial, expert, easily accessible and capable of issuing decisions in a "swift, efficient and cost-effective manner" (Art. 10(1)–(2), p. 45). These institutional provisions embed routes for recourse and independent review.

Trust

Stakeholders consider *Trust* crucial, emphasizing that the Data Act should ensure both high data quality and meaningful transparency. They argue that effective data sharing and subsequent use by related technologies require robust data-quality standards: *"high standards of data quality are essential for decision-making, providing efficient and reliable services while enhancing public trust"* [F2669460]. At the same time, they recognise that enforcing such standards can impose substantial administrative burdens, since data often need to be cleaned and processed before they can be shared: *"There are many practical issues with data sharing, such as the need to clean and process the data before it is released, which means it is burdensome for companies to release it"* [F2669299]. Trust is therefore framed as a balance between reliability for users and feasibility for organisations.

To address these practical challenges, stakeholders propose participatory and calibrated transparency mechanisms. In domains such as healthcare, they emphasise that trustworthy exchanges depend on inclusive decision-making: *"active participation of a broad range of stakeholders, including healthcare professionals, medical societies and patients, must be ensured in taking decisions around trustworthy data exchanges in the EU"* [F2669426]. Transparency obligations, in their view, must simultaneously safeguard legitimate interests: *"transparency obligations need to ensure that data holders' legitimate interests are duly protected"* [F2669492]. Several letters therefore recommend targeted disclosures about data activities and outcomes, except where this would conflict with the public interest, as a way to enhance accountability and public trust: *"disclose the data activities performed on the data and the derived results, unless such disclosure would be contrary to the public interest. This would not only ensure the accountability of the public body but also increase trust"* [F2669492]. More broadly, contributors suggest that trust conditions should be written into contractual arrangements, including guarantees on the nature of the data, hosting arrangements, reuse intentions and limits, transparency duties, and clear allocation of managerial responsibilities: *"The conditions of this trust could be the following: obtaining guarantees on the nature of the data shared, their hosting, the intention behind their reuse, the conditions and possible limits to this reuse, or even obligations of transparency of processing or clear rules relating to the responsibilities of the manager"* [F2669275].

The EU Data Act likewise embeds trust through contractual clarity, confidentiality protections, and explicit transparency and traceability duties. It requires contracts between businesses to specify obligations and liability related to data access, processing, sharing, and storage, so that responsibilities are clear and misuse can be addressed. It also provides safeguards for confidential business information and trade secrets, recognising them as essential to the functioning of the internal market. In addition, the Act mandates that data holders inform users how user-generated data will be used, and introduces traceability requirements whereby platforms must describe to business users the data generated through their services. These combined provisions aim to make data practices understandable and predictable for affected parties, while protecting sensitive commercial information.

Ownership and Property

Stakeholders articulate the value of *Ownership and Property* primarily in terms of intellectual property rights (IPR) and contractual control across their data ecosystems. They portray data as a valuable and reusable resource, but caution against exclusive property rights over data itself: *"exclusive property rights are not the right way to regulate the use of data [...] the discussion should not be about property rights, but about rights of use to which several people can be entitled in parallel"* [F2669271]. To safeguard data assets from unauthorised access and use by external actors, they highlight contractual mechanisms such as *"contracting with third parties or enforcement against unauthorised use"* [F2669284]. Several letters stress that the capability to process data constitutes a form of intellectual capital that underpins monetisation potential and therefore requires protection at each stage: *"For companies, the ownership of data must be regulated by contract. Data processing implies de-facto intellectual property and the latter must be protected with patent or copyright laws, to become an Intellectual Property Right. Companies must be able to protect each stage through contracts"* [F2669212]. At the same time, stakeholders warn against overly uniform regulatory solutions and argue that restrictions on data assets should reflect sectoral specificities rather than follow a *"'one size fits all' approach that does not meet the needs of businesses and/or their vendors"* [F2669216].

The EU Data Act also engages with ownership and property, but does so by focusing on user control, fair competition, and balanced compensation rather than by introducing new exclusive data rights. A central element is strengthened user control over data generated by the use of products and related services, designed to allow users to access and share such data with alternative service providers. The Act explicitly links this to competition in aftermarket services, stating that consumers should "benefit from a wider choice in aftermarket services, such as repair and maintenance, and no longer depend on only the manufacturer's services". It further sets out principles for compensating data holders when they are legally required to make data available, thereby preserving incentives to invest in data-generating infrastructures, but it caps compensation for micro, small, and medium-sized enterprises so that it does not exceed direct costs, protecting them from undue economic burden. In situations of exceptional public interest, such as public emergencies, the Act clarifies that the public interest in using the data can outweigh the data holder's freedom to dispose of the data, thus incorporating a limited public-interest override into the ownership landscape.

Environmental sustainability

Stakeholders articulate the value of *Environmental Sustainability* primarily in terms of supporting economic, data, and environmental ecosystems through collaborative data use. They strongly advocate cooperation between private and public sectors on data sharing, viewing this as critical for tackling broader societal objectives such as climate change and financial crime [F2669435]. In their view, the Data Act should bolster economic sustainability by enabling innovation and data-driven value creation, emphasising the importance of "unlocking innovation and value through big data, artificial intelligence (AI) and other emerging technologies" [F2669282]. A robust data ecosystem with broad access is portrayed as foundational for growth, innovation, and regional competitiveness [F2669282].

At the same time, stakeholders caution that extensive data sharing must be designed to avoid undermining competition and innovation. They argue that regulatory measures should "avoid adverse effects on competition and innovation" [F2669270] and stress that sustaining the data ecosystem via multi-actor collaboration is necessary to "unlock the full potential of data and innovative technologies" [F2669282]. By promoting comprehensive data availability and exchange, they anticipate enhanced regional business capacities and smoother adaptation, arguing that "the better we exchange data, the less disruptive this will be to current business models and organisations" [F2669244]. Stakeholders also highlight the direct environmental role of data: aggregated data can support responses to epidemics, natural disasters, and pollution [F2669270], and enhanced data-driven services can help "consumers take action to reduce their environmental impact" and encourage "individual action to reduce the environmental footprint" [F2669459].

The EU Data Act similarly embeds environmental sustainability in its framing of data as a resource for public-interest goals. In line with the European Green Deal, its recitals emphasise mobilising private-sector data to address climate change, biodiversity loss, pollution, and the sustainable management of natural resources, and to strengthen preparedness, response, and recovery in environmental and other public emergencies (Recitals 6, 57). The Act's provisions on business-to-government access under conditions of exceptional need are explicitly justified in terms of enabling public bodies to use privately held data to manage crises and support resilience in areas such as health, disaster response, and environmental protection (Chapter V). Taken together, these elements position environmental sustainability as a value that is operationalised through data access for public-interest purposes, crisis management, and long-term ecological objectives.

3.4 Discussion

3.4.1 From Values to Design: What the EU Case Reveals About Data Democratization

The Phase I and Phase II results, the five VSD values do not appear as isolated concerns, but as recurring constellations around a small set of design problems. Across stakeholder letters and the EU Data Act, *Universal Usability* and *Ownership and Property* cluster around questions of who can access and use data, on what terms, and with which protections for weaker actors. *Accountability* and *Trust* repeatedly surface when responsibilities, safeguards, and recourse need to be made clear and usable in practice. *Universal Usability*, *Accountability*, and *Environmental Sustainability* intersect where data flows are expected to support broader public interest objectives, including fair competition, resilience, and environmental goals. Finally, *Universal Usability* and *Trust* link to portability and interoperability wherever nominal rights must be translated into workable migration and switching paths.

Reading these value clusters alongside the coded excerpts and the relevant legislative mechanisms, we distilled a minimal and non-overlapping set of higher-order themes that make the value configurations actionable for policy design. They are *equitable access* (rights that weaker actors can actually exercise), *protected openness* (sharing under clear safeguards for legitimate interests), *pragmatic interoperability* (infrastructure and procedures that make access usable), and *participatory accountability* (obligations and remedies that are legible and enforceable for non dominant actors). These meta requirements synthesise the value level insights into mechanism oriented design logics that we then interpret through the lens of data democratization and existing IS scholarship.

These meta-requirements are consistent with, and are grounded in, established streams of IS and policy scholarship. *Equitable access* integrates universal usability, autonomy, and ownership where the core problem is who gains access to which data and under what conditions, aligning with data governance work on decision rights and participation (Khatri and Brown, 2010; Abraham et al., 2019) and with organizational studies of data democratization that emphasize widening access and capability beyond expert elites (Lefebvre et al., 2021; Gregory et al., 2018). *Protected openness* links sharing with safeguards by bundling privacy, informed consent, and freedom from bias, which reflects canonical VSD commitments and cumulative IS privacy research (Friedman et al., 2013, 2019; Smith et al., 2011; Malhotra et al., 2004) and work that specifies how transparency becomes enforceable accountability in socio-technical systems (Williams et al., 2022). *Pragmatic interoperability* surfaces where accountability co-occurs with usability and ownership in switching and standards contexts and is supported by research on portability and interoperability as instruments for lowering structural frictions and enabling value creation across ecosystems (Hodapp and Hanelt, 2022; Kim, 2025; Rubinfeld, 2023). Finally, *participatory accountability* links accountability, trust, and consent-related transparency where responsibilities must be legible and enforceable for weaker parties, connecting VSD's accountability focus with IS work on explanation, traceability, and recourse in high-stakes decision environments (Fernández-Loría et al., 2022; Schulze et al., 2023; Kronblad et al., 2024).

Meta Requirement 1: Equitable access

Our first meta requirement, equitable access, captures the idea that democratization is not achieved by declaring rights in the abstract, but by ensuring that weaker actors can in practice obtain and use data under workable conditions. Across the stakeholder letters, universal usability and ownership concerns converged on calls for a "level playing field" in which SMEs and non-dominant actors can share in the benefits of data-driven innovation rather than being locked into the infrastructures and terms of a few dominant manufacturers. In the Data Act, this concern materializes in user-facing access rights, obligations on manufacturers to make data generated by products easily retrievable, and capped compensation regimes for micro and small enterprises. Read together, the findings show that equitable access is understood as a coupling of entitlement, affordability, and basic literacy, not as a purely formal grant.

This interpretation resonates with IS governance work that treats data governance as a problem of allocating decision and access rights across actors (Khatri and Brown, 2010; Abraham et al., 2019). Governance arrangements that concentrate decision rights in a narrow elite may formally preserve access for others, but they do not expand meaningful participation. Organizational research on data democratization makes a similar point, showing that making data broadly available only affects behavior when organizations also build capabilities such as self-service analytics, data literacy, and supporting structures for non-expert users (Lefebvre et al., 2021; Gregory et al., 2018). Equitable access in our case aligns with this capability-oriented view, that is, rights that travel are rights backed by the means to exercise them.

Seen through this lens, the EU design offers an example of how law can operationalize equitable access as entitlement plus capability. By tying user access to explicit duties on providers, by recognizing provision costs and capping them for smaller firms, and by situating access within a broader program of interoperability and switching, the Act moves closer to what capability theorists describe as the difference between having a right and having the substantive freedom to act on it (Sen, 1999; Nussbaum, 2011). For data democratization, the implication is that policy needs to take the "last mile" seriously. It is not enough to say that SMEs can request data, the regime must make such requests intelligible, financially viable, and compatible with the technical realities of their operations.

Meta Requirement 2: Protected openness

The second meta requirement, protected openness, reflects the recurrent pattern in our findings where stakeholders accept wider data sharing, including cross-border and business-to-government flows, on the condition that legitimate interests are safeguarded. Stakeholders framed openness as desirable when it is grounded in calibrated transparency, voluntary or consented arrangements, and risk-sensitive terms that take account of the sensitivity of data and the costs of preparing it. The Act responds with a dense layer of safeguards, which is protections for trade secrets and confidential business information, clarification of the relationship with database rights, user-requested third-party access as the default, and narrowly framed conditions under which public-interest claims can override the data holder's preferences. Together, these mechanisms depict openness as conditional rather than unconditional.

This configuration mirrors long-standing insights from IS privacy research. Classic models such as the Concern for Information Privacy and IUIPC scales conceptualize privacy not as secrecy, but as control mediated by awareness of practices, consent, and institutional guarantees (Smith et al., 2011; Malhotra et al., 2004). Value Sensitive Design similarly treats privacy, informed consent, and freedom from bias as core values that must be embedded into socio-technical systems through design choices (Friedman et al., 2013, 2019). More recent work on transparency and accountability argues that transparency only becomes effective when it is tied to mechanisms for contestation and redress, rather than left as information disclosure alone (Williams et al., 2022). Our findings show the same logic at policy scale: stakeholders do not simply ask for "more transparency", they ask for transparency that is proportionate, actionable, and compatible with confidentiality obligations; the Act codifies this demand in concrete rights, exceptions, and technical safeguards.

For data democratization, protected openness is therefore the boundary condition that prevents sharing from collapsing into extraction. A democratized regime cannot be equated with maximal openness, because unconstrained access tends to advantage those with greater processing and bargaining capacity. Instead, the EU case suggests that democratization depends on specifying who can request what data, under which purposes, with which protections for trade secrets and privacy, and with what forms of compensation. This meta requirement connects democratization to the institutionalization of guardrails, making clear that "open by default" must be matched with "protected by design" if weaker actors are not to pay the price for wider data flows.

Meta Requirement 3: Pragmatic interoperability

The third meta requirement, pragmatic interoperability, captures the work needed to turn nominal access into usable capability. In the empirical material, stakeholders repeatedly describe portability and interoperability not as ends in themselves, but as enabling conditions for exercising rights and participating in data ecosystems. They stress that technical standards must be tailored to sectoral realities, that portability can be infeasible for certain data types, and that top-down or incumbent-driven standards risk excluding smaller firms. The Act, in turn, treats interoperability and switching as deliberate market design. It bundles portability rights with provider-switching duties, limits on switching charges, pre-contractual information obligations, and references to common specifications and open interoperability standards.

In IS research, interoperability has long been framed as an infrastructural condition for value creation across digital ecosystems, rather than a mere technical feature (Hodapp and Hanelt, 2022). Work on data portability and switching shows that reducing structural frictions lowers dependency on dominant platforms and allows complementors and SMEs to reconfigure their participation in ecosystems (Rubinfeld, 2023). At the same time, governance scholarship warns that poorly calibrated standards can entrench incumbents and create new forms of lock-in if standard-setting processes are captured by dominant actors (Markus et al., 2006a). Our findings thus echo the following concerns, stakeholders welcome interoperability as an enabler of participation, but insist on sector-sensitive, community-led standardization to avoid reinforcing existing power asymmetries.

Pragmatic interoperability, as it emerges from the EU case, is therefore more than the presence of APIs or format specifications. It is a mechanism bundle that includes technical standards, switching rights, transparency about exit conditions, and harmonized procedures for cross-border flows. For democratization, this means that regulators interested in widening access need to treat interoperability as infrastructure: something that must be financed, governed, and periodically recalibrated, not merely encouraged. Without this layer, access rights remain brittle; with it, rights can be exercised at reasonable cost and at the speed required by contemporary digital operations.

Meta Requirement 4: Participatory accountability

The fourth meta requirement, participatory accountability, integrates value signals around accountability, trust, and informed consent where the central issue is whether non-dominant actors can understand, contest, and rely on the arrangements that govern data sharing. Stakeholders consistently ask for harmonized procedures rather than new obligations, for clear role definitions between users, vendors, and intermediaries, and for contracts that spell out responsibilities, risks, and remedies in a way that is usable by SMEs. They also link trust to feasible data-quality expectations, stakeholder involvement in sensitive domains, and calibrated transparency that does not unduly expose trade secrets. The Act responds with unfair-terms controls in B2B data-sharing contracts, pre-contractual information duties, requirements for smart-contract safety, and safeguards around trade secrets and business confidentiality, alongside conditions that limit business-to-government requests to situations of exceptional need.

These mechanisms resonate with IS governance work that treats accountability as a matter of traceability and answerability across distributed socio-technical systems (Khatri and Brown, 2010). Algorithm and platform accountability research similarly emphasizes the importance of legible obligations, logging and auditability, and accessible remedies for those affected by automated decisions (Cobbe et al., 2023; Rai, 2020). Value Sensitive Design explicitly includes accountability and trust as values that must be embedded through design features that allow stakeholders to attribute responsibility and seek redress (Friedman et al., 2019). Our meta requirement extends these insights to a policy context: participatory accountability is the point where duties, contracts, and technical instruments such as smart contracts must be configured so that weaker actors can realistically invoke their rights and challenge unfair terms.

For data democratization, the lesson is straightforward but demanding. Participation without recourse is not participation. A regime that expands access but leaves accountability opaque will likely reproduce existing hierarchies, as only well-resourced actors can navigate the resulting complexity. The EU case illustrates an alternative: match expanded sharing with mechanisms that make obligations auditable and remedies reachable at reasonable cost. That requires attention not only to formal rules, but also to the cognitive and financial load imposed on SMEs when they try to understand international access risks, negotiate contract terms, or respond to public-interest data requests. Participatory accountability, in this sense, is the glue that holds the other three meta requirements together: equitable access, protected openness, and pragmatic interoperability all rely on accountability structures that less powerful actors can actually use.

Read as a set, these meta-requirements translate value talk into policy-relevant design problems and mechanisms. Conceptually, they align with data democratization understood as broadening the circle of legitimate data users while preserving rights, capabilities, and recourse for those with fewer resources. Practically, the EU Data Act case illustrates an approach in which user entitlements and fair allocation (MR1) are coupled to calibrated safeguards (MR2), then rendered operable through switching and interoperability (MR3), and finally anchored in contractual and institutional arrangements that make obligations usable for weaker parties (MR4). In other words, democratization emerges not from a single lever but from a sequenced bundle; who gets access, on what protected terms, by which operational means, and with what recourse

3.4.2 Lessons About Operationalizing Democratization Learned by the EU Design

Building on the four meta-requirements, we now examine how the EU design makes them concrete. Rather than treating the Data Act as a stand-alone piece of legislation, this section interprets how the Data Act assembles specific mechanisms around each meta requirement and how these mechanisms speak to the practical challenges of data democratization. Section 5.3 then generalizes from this case reading into broader design lessons and policy propositions.

Equitable access as entitlement plus capability.

The EU Data Act moves beyond declarative access by coupling user entitlements with mechanisms that make access exercisable in practice. Stakeholders insisted that arrangements should "take into account" SMEs and "actively avoid and rectify" imbalances. The Act's user-facing information duties and capped compensation for smaller firms instantiate that request in legal form. This pairing mirrors the capability perspective of data democratization that distinguishes formal rights from the substantive freedoms and means to use them (Sen, 1999; Nussbaum, 2011). In IS governance, allocating decision and access rights is necessary but insufficient unless rights are paired with affordability, intelligibility, and process clarity that widen effective participation (Khatri and Brown, 2010; Abraham et al., 2019). Organizational research on data democratization similarly shows that broader access must be supported by self-service analytics, literacy, and role clarity if non-elite actors are to generate value rather than remain nominal rights-holders (Lefebvre et al., 2021; Gregory et al., 2018). The EU Data Act case operationalizes this logic, that is, information duties lower interpretive barriers, compensation caps reduce financial barriers for SMEs, and aftermarket provisions translate access into competitive choice. Equitable access in this context is therefore not just a statement of principle but a carefully engineered linkage between rights, capabilities, and market structure.

Protected Openness as the Boundary Condition for Sharing

Stakeholders endorsed openness when legitimate interests are safeguarded, asking for calibrated transparency, contractual vehicles, and risk-sensitive cross-border pathways. The legislation answers with trade-secret protections, limits on database-rights interference, and user-requested third-party access which matches long-standing insights that transparency alone does not yield accountability without institutions that specify audiences, obligations, and remedies (Fox, 2007; Bovens, 2007). Privacy scholarship likewise cautions that sharing must respect contextual norms and role-based expectations rather than rely on one-size disclosure (Nissenbaum, 2004, 2010), and IS work shows that durable privacy concern models translate into enforceable safeguards only when embedded in concrete governance mechanisms (Smith et al., 2011; Malhotra et al., 2004). Recent critiques of "transparency-only" approaches underline the need to link visibility to actionability and recourse (Ananny and Crawford, 2018). The EU design follows this logic by specifying when, how, and to whom sharing occurs, under what proportional terms, and with what carve-outs for trade secrets and exceptional-need access. As a design principle, protected openness separates the value of sharing from the risk of extraction by binding openness to safeguards ex ante.

Pragmatic Interoperability as the Work That Makes Access Usable.

Stakeholders treated portability and interoperability as means rather than ends, warning against one-size standards and incumbent-led coordination, thereby, the Act adds switching provisions and standardization pathways that reduce structural frictions. Economic theory predicts that portability and interoperability lower switching costs and weaken lock-in, increasing contestability and user surplus (Farrell and Klemperer, 2007; Shapiro and Varian, 1999). In digital ecosystems, interoperability is infrastructural: it enables complementary innovation across modular layers and platforms (de Reuver et al., 2018; Baldwin and Clark, 2000). IS work similarly frames interoperability as a condition for value creation across data ecosystems rather than a technical afterthought. The EU bundle takes this seriously by combining user portability with provider switching duties and references to common specifications. For democratization, the implication is that access must be rendered operable through migration pathways and conformance regimes that are sector-tuned and community-led, so smaller firms can plug in without bearing the costs of bespoke integration or the risks of incumbent-designed standards.

Participatory Accountability to Keep Burdens Proportionate.

Both the stakeholder letters and the Data Act locate accountability in contracts, information duties, and fairness controls that non-dominant actors can actually use, not just in formal allocations of responsibility. Public administration research emphasizes that accountability requires answerability and enforceability, which in practice hinge on clarity of obligations, visibility of actions, and accessible remedies (Bovens, 2007; Fox, 2007). VSD's focus on integrating ethical values into socio-technical artifacts points to concrete design levers such as traceability and explainability that make obligations auditable (Friedman et al., 2019). Work on algorithmic accountability similarly argues that disclosures and explanations must be actionable for affected parties to contest or seek remedy (Doshi-Velez and Kim, 2017; Ananny and Crawford, 2018). The EU case advances these ideas via unfair-terms controls in B2B data sharing, pre-contractual risk disclosures, traceability duties, and essential requirements for automated execution. For democratization, the lesson is simple; participation without recourse is not participation. Accountability needs to be proportionate, legible at handoff points, and enforceable at reasonable cost, particularly for SMEs.

3.4.3 Design Lessons and Policy Propositions from the EU Data Act Case

Building on the four meta-requirements identified in our Findings, we distill four actionable lessons for policy design when the goal is data democratization:

Engineer Rights as Capabilities

Access provisions are effective only when paired with the capabilities that make them usable by non-dominant actors. The capability perspective distinguishes nominal rights from the substantive means to exercise them (Nussbaum, 2011). In IS governance, allocating decision and access rights must be complemented by affordability, intelligibility, and role clarity to widen meaningful participation (Khatri and Brown, 2010; Abraham et al., 2019). Organizational work shows that broadening access without supporting self-service use and literacy seldom changes behavior (Lefebvre et al., 2021). Institution-based trust research further indicates that formal guarantees and clear information duties reduce perceived risk and enable uptake by weaker parties, which aligns with our finding that stakeholders sought fairness for SMEs and the Act coupled user information duties with capped compensation and aftermarket choice (Pavlou and Gefen, 2004).

Policy proposition: Pair access rights with simple guidance, low and predictable costs, and basic support so smaller actors can actually use their data, and judge success by real uptake and timely switching rather than formal compliance.

Open by Design, Protected by Default.

Sharing at scale requires front-loaded safeguards that translate visibility into accountability and respect contextual privacy norms. Transparency alone does not self-execute into accountability without institutions that specify audiences, obligations, and remedies (Bovens, 2007; Fox, 2007), and IS privacy research emphasizes contextual integrity and enforceable controls over disclosure-only approaches (Bélanger and Crossler, 2011; Smith et al., 2011; Malhotra et al., 2004; Nissenbaum, 2004). This logic is reflected in the case where stakeholders endorsed calibrated transparency, voluntary contractual templates, and risk-sensitive cross-border pathways, and the Act protected trade secrets, clarified database rights, and conditioned third-party access on user request.

Policy proposition: Make sharing user-led and governed by standard protective clauses that define confidentiality and redress, and assess balance by fewer disputes and faster completion of public-interest requests.

Treat Interoperability as Infrastructure.

Interoperability turns formal access into practical use by lowering switching costs and weakening lock-in (Farrell and Klemperer, 2007; Shapiro and Varian, 1999). In digital ecosystems it is infrastructural, enabling complementor participation and distributed innovation (de Reuver et al., 2018). IS work on standards shows that sector-tuned, open, and participatory processes are less likely to encode incumbent bias and are more likely to diffuse (Markus et al., 2006b; Lyytinen and King, 2006; Hanseth and Lyytinen, 2010). Consistent with our results, stakeholders argued for community-led specifications and harmonised procedures, and the Act targeted lock-in while underlining portability, switching, and common specifications.

Policy proposition: Design portability, provider switching, and common specifications as one package, and test effectiveness by shorter time-to-exit, higher SME conformance on open test suites, and fewer exclusive technical dependencies in new contracts.

Make Accountability Usable for the Least Powerful.

Accountability becomes participatory when obligations are legible and remedies are accessible at reasonable cost, which requires explanations and traceability that affected parties can act upon (Bovens, 2007; Gregor and Benbasat, 1999). Institution-based trust findings show that formal controls and third-party assurances reduce transaction risk for weaker parties (Pavlou and Gefen, 2004), and our case shows stakeholders requested harmonised procedures, role clarity, and workable notice while the Act added unfair-terms controls, pre-contractual disclosures, traceability, and essential requirements for automated execution.

Policy proposition: Combine unfair-terms controls with standard risk disclosures and minimum auditability for automated execution, and gauge performance by improved SME dispute outcomes and faster time-to-remedy.

These lessons reflect how the EU Data Act case operationalizes democratization without introducing new assumptions beyond the evidence we analyzed. They also generalize as design logics for other jurisdictions where similar tensions appear.

3.4.4 Implications for organizations and SMEs

A democratized data regime shifts advantage from exclusive possession of raw data to the capabilities that make data reusable, portable, and governable in practice. In IS terms, performance follows from bundles of governance, quality, architectural, and contractual capabilities rather than from stockpiles of records (Khatri and Brown, 2010). First, firms will need to treat data quality as fitness for use across contexts, and not merely accuracy in source systems (Wang, 1996; Marchand et al., 2000). Provenance, cataloging, and consent management become routine controls that enable lawful reuse and reduce coordination costs across partners (Bélanger and Crossler, 2011; Smith et al., 2011; Nissenbaum, 2010). Second, explainability and traceability move from compliance add-ons to operating requirements because they enable answerability and remedy when value tensions surface (Gregor and Benbasat, 1999; Doshi-Velez and Kim, 2017). Third, architectural choices that expose boundary resources and standard interfaces increase option value, since complementors and customers can connect without bespoke integration (Ghazawneh and Henfridsson, 2013; de Reuver et al., 2018; Hanseth and Lyytinen, 2010). Finally, contractual competence matters, because standardized clauses and fair-terms controls reduce transaction risk and support participation by weaker parties (Pavlou and Gefen, 2004).

For SMEs, the EU Data Act case indicates that lowered information burdens and capped compensation reduce entry costs into data collaborations, yet benefits materialize only if organizations have absorptive capacity to act on accessible data. IS work on data democratization highlights that self-service analytics, data literacy, and collaborative culture are necessary complements to access, otherwise rights remain inert (Lefebvre et al., 2021). SMEs should therefore prioritize lightweight data catalogs, template-based contracts, and participation in sector association that co-produce workable specifications, which aligns with evidence that standards diffuse through collective action rather than top-down decree (Markus et al., 2006b; Lyytinen and King, 2006). Vendor management should emphasize reversible commitments and exit readiness, since portability and switching reduce exposure to lock-in and support bargaining power in renegotiations (Farrell and Klemperer, 2007; Shapiro and Varian, 1999).

For incumbents, advantage is likely to pivot toward orchestration and reliability rather than exclusivity. Interoperability-friendly architectures, documented exit paths, and auditable logs anticipate user-requested sharing and public-interest access under defined conditions, which lowers dispute risk and sustains trust with regulators and partners. Platform governance that balances control and contribution, for example through stable boundary resources and transparent terms, sustains third-party innovation while meeting fairness and accountability requirements (Ghazawneh and Henfridsson, 2013). In short, organizations that invest in quality, lawful reuse, explainable processing, switching readiness, and standard-conformant interfaces are better positioned to capture value in democratized data ecosystems where participation, not possession, is the primary constraint.

3.4.5 Situating The EU Approach Among Major Economies

European Union The EU advances democratization as *rights plus safeguards plus operability* in a single horizontal instrument. The Data Act links user access to data generated by connected products and related services with enforceable portability, provider switching, and interoperability duties for data processing services, while protecting trade secrets, clarifying the non-application of database sui generis rights in this context, and conditioning third-party access on user request (European Union, 2023). These mechanisms sit alongside the Data Governance Act's trust infrastructure for reuse of protected public-sector data, data intermediation, and data altruism schemes (European Union, 2022b). The result is a regulated commons where equitable access is instantiated through user-facing entitlements and capped compensation for smaller firms, protected openness is specified in law, and pragmatic interoperability is engineered as market design rather than left to ad hoc coordination (European Union, 2023, 2022b).

United States. The United States continues to rely on sectoral statutes and market-led coordination, which yields uneven progress on portability and interoperability across domains. Health information illustrates how targeted federal rule-making can operationalize portability at scale via APIs and anti–information blocking obligations under the 21st Century Cures Act and the ONC Health IT Certification Program (Department of Health and Human Services, 2020). Consumer privacy rights, including data access and portability, have expanded at the state level through the CCPA as amended by the CPRA rather than through a comprehensive federal privacy statute (California Consumer Privacy Act, 2023). In this environment, our four meta-requirements translate into complementary levers: unfair-terms and disclosure controls under consumer-protection authority to support participatory accountability, sector-specific portability to advance equitable access, procurement and standards-body work to deliver pragmatic interoperability, and privacy-anchored accountability regimes to sustain protected openness. The actionable emphasis, relative to the EU, is on stitching together rights and capabilities across agencies and sectors rather than enacting a single horizontal framework.

China. China's data regime foregrounds national security, data sovereignty, and platform governance through the Cybersecurity Law, the Data Security Law, and the Personal Information Protection Law, with localization and structured cross-border transfer pathways shaped by security assessments, certifications, and contractual safeguards (Latham and Watkins, 2025; Tan and Gu, 2025). Recent regulatory adjustments have eased some cross-border flows while clarifying categories and procedures, yet oversight remains centralized and risk-tiered (Reuters, 2024a,b). Within our lens, protected openness is defined predominantly by state priorities, participatory accountability is mediated by licensing and supervisory controls, and democratization is orchestrated within designated domains rather than generalized through market-wide switching and interoperability.

Cross-jurisdictional takeaway. The EU demonstrates a coherent blueprint that bundles entitlements, guardrails, and operability in one statute. The United States shows a path that advances similar aims through sectoral rulemaking, state privacy rights, and standards-driven procurement. China illustrates a sovereignty-first model that conditions access and sharing on security classifications and export controls. Read through our meta-requirements, where equitable access is limited to nominal rights, participation concentrates among incumbents; where protected openness lacks enforceable guardrails, sharing either stalls or erodes legitimate interests; where pragmatic interoperability is thin, access remains inert; and where participatory accountability is not usable by weaker parties, rights are difficult to exercise in practice. These lenses provide an actionable diagnostic for incremental reform even without wholesale legislative overhaul.

3.4.6 Implications for theory

This study advances IS theorizing about data governance in three ways. First, this study offers a mechanism-based account of how data democratization is instantiated in practice. Rather than treating participation as the outcome of any single lever, our analysis specifies a configuration of four mechanism families that together make participation possible: rights that translate into usable capabilities, safeguards that condition sharing, interoperability engineered as shared infrastructure, and accountability that is legible and enforceable for the least powerful. This specification complements organizational perspectives that emphasize access, literacy, and culture by adding policy-level design logics that can be observed and measured in law and contracts. The theoretical pay-off is a set of mid-range mechanisms that identify where participation is likely to fail and that support falsifiable propositions about necessary and sufficient conditions for broadened access.

Second, by treating legislation as a governance artifact, the study extends Value Sensitive Design into techno-regulatory design. The meta-requirements show how value constellations become implementable instruments such as entitlements, exceptions, standards, and contract controls, thereby clarifying the translation from value identification to institutional embedding. This moves VSD beyond the design of products and services toward the design of rules that shape socio-technical environments, while preserving VSD's focus on identifying, analyzing, and reconciling value tensions in context.

Third, the analysis links classic IS governance to ecosystem-level governance of data. Decision rights and accountability have traditionally been theorized inside firms; our findings locate their analogs in ecosystem mechanisms that require portability-ready architectures, boundary resources for third-party integration, contract portfolios aligned with unfair-terms regimes, and transparency programs that satisfy pre-contractual information duties without exposing trade secrets. The contribution, hence, is a bridge from intra-organizational governance to shared market infrastructure, which supports theorizing about capability complementarities between organizations and regulators and about their joint effects on participation and competition.

3.4.7 Limitations and Future Works

The claims in this paper are bounded by a single horizontal regime that targets non-personal data sharing and by a corpus consisting of stakeholder consultation letters and enacted legislative text. External validity is strongest for jurisdictions that pursue pro-competitive access and reuse through legal entitlements and standardization. Several limitations follow. The quantitative value mapping operationalizes VSD categories through supervised NLP; construct validity depends on the adequacy of these operationalizations and thresholding choices. The qualitative synthesis compares letters and provisions rather than postenactment behavior, so we do not claim causal effects on market outcomes. Stakeholder heterogeneity is analyzed thematically rather than exhaustively stratified by sector or firm size. Text availability and selection may bias which arguments are represented in the consultation record. Legal interpretation and enforcement can vary across Member States and over time as guidance, delegated acts, or standards evolve. Future research can examine portability of the mechanism bundle across sectors and jurisdictions through comparative case designs, and estimate the marginal effects of specific instruments using quasi-experimental strategies. Measurement can move beyond text to market traces and administrative data, for example the rate and timing of portability requests, switching events, dispute outcomes for SMEs, and the adoption of standardized clauses. Organizational studies can investigate capability reconfiguration inside firms that respond to rights, safeguards, and operability obligations, and how these capabilities interact with sector consortia and standards bodies. Together, these extensions would test where each mechanism is binding, where complementarities arise, and how democratization diffuses over time.

Conclusion

This dissertation set out to address a practical and theoretical tension in IS research. ML offers powerful capabilities for prediction and pattern discovery, yet ML systems often face barriers to legitimate and effective use in socio-technical settings because they can be opaque, demanding in data and implementation requirements, and misaligned with institutional values. Across three studies, the dissertation argues that these challenges are best understood as socio-technical design problems rather than as purely technical shortcomings. The central claim is that ML in IS should be treated as socio-technical decision infrastructure, an artifact that participates in decision processes and therefore must be evaluated by criteria that extend beyond accuracy to include interpretability for stakeholders, feasibility under real constraints, and value sensitivity (Padmanabhan et al., 2022; Selbst et al., 2019).

The three chapters address different domains and methods, but they are unified by a common design logic as each constructs an ML based artifact intended to support decision-making under a specific socio-technical constraint.

Chapter 1 focuses on *operational feasibility under measurement constraints*. It shows that information acquisition can be treated as a resource allocation problem and that reinforcement learning can be used to reduce measurement burden while preserving decision quality (Janisch et al., 2020; Sutton and Barto, 2018). The DRL agent learns which survey items provide the greatest discriminative value, enabling practitioners to obtain reliable maturity classifications without administering lengthy assessments, a critical requirement for deployment in schools and youth services.

Chapter 2 focuses on *epistemic accountability in ML-enabled research*. By coupling interpretable ML discovery with confirmatory statistical testing, it demonstrates a disciplined workflow that leverages ML's flexibility without collapsing inference into pattern mining (Shmueli, 2010; Lundberg and Lee, 2017). The discovery-confirmation workflow ensures that the relationships identified by SHAP are not merely descriptive correlations but can withstand inferential scrutiny, strengthening their value for developmental theory and intervention design.

Chapter 3 focuses on *normative complexity in governance*. By mapping stakeholder value emphases and linking them to implementable regulatory mechanisms, it shows how ML can support value-sensitive analysis of techno-regulatory design (Friedman et al., 2006, 2019; European Union, 2023). The NLP-based analysis surfaces tensions that might otherwise remain implicit in policy discourse and translates them into meta-requirements that can inform regulatory implementation.

These studies together support a higher-level synthesis that responsible ML in IS requires aligning model design with the institutional conditions of use. The artifact is not merely the predictive model, it is the full socio-technical arrangement that includes what is measured, how outputs are generated, how they are explained, and how they translate into action under governance constraints. Three integrative contributions arise from these individual studies:

Interpretability as a Bridge from Outputs to Action: A first contribution is to treat interpretability as a functional bridge between model outputs and institutional action. In the digital maturity studies, interpretability supports intervention-relevant reasoning by clarifying which competencies drive classifications and how relationships change across developmental stages (Doshi-Velez and Kim, 2017; Lundberg and Lee, 2017). In the governance study, interpretability supports traceability from computational patterns in text to concrete policy mechanisms, enabling value claims to be examined as design commitments rather than accepted as opaque outputs. This aligns with the view that explanations should be evaluated relative to purpose and audience, and with cautions that post-hoc explanation alone does not guarantee accountability (Rudin, 2019).

Feasibility as Part of Performance: A second contribution is to conceptualize feasibility as part of performance rather than as a downstream implementation concern. Many organizational settings cannot sustain data collection practices that are costly in time, attention, or compliance burden. Chapter 1 shows that designing cost-aware decision support artifacts can make the difference between an ML approach that is theoretically impressive and one that is practically deployable (Janisch et al., 2020; Sutton and Barto, 2018). More broadly, the dissertation suggests that socio-technical ML artifacts should be evaluated under realistic constraints on measurement and institutional capacity, particularly in settings such as education, public services, or healthcare where resources and legitimacy are bounded.

Values as Design Variables Linked to Mechanisms: A third contribution is to demonstrate how values can be treated as design variables rather than external commentary. Values enter ML systems through choices about objectives, proxies, thresholds, and governance arrangements (Selbst et al., 2019; Mitchell et al., 2021). Chapter 3 extends this logic to policy by linking stakeholder value emphases to implementable governance mechanisms and by deriving meta-requirements that translate data democratization goals into designable institutional commitments (Friedman et al., 2006, 2019; European Union, 2023). This offers a structured approach for IS research to connect normative analysis to concrete socio-technical mechanisms.

ML Artifacts and Socio-Technical Feedback

A crucial characteristic of ML systems in IS contexts is their participation in feedback loops. The predictions generated by ML do not merely describe reality, they intervene in it. A digital maturity assessment, once delivered to an educator, shapes the interventions that student receives, which in turn affects their future digital behavior and subsequent assessments. Similarly, a policy framework designed using value-sensitive NLP analysis will generate stakeholder responses that reshape the next policy iteration. This reflexivity, where ML outputs become inputs to the social systems they model, positions ML as socio-technical infrastructure rather than a passive measurement tool, recognizing this feedback dynamic has several implications. First, it suggests that ML artifacts require continuous governance rather than one-time validation. Second, it implies that socio-technical design principles must be embedded from inception rather than retrofitted after deployment. Third, it motivates a forward-looking research agenda that studies how ML artifacts behave after deployment, how stakeholders adapt to them, and how technical and institutional controls co-evolve over time (Faraj et al., 2018; Berente et al., 2021). Unlike traditional IT artifacts that remain relatively stable once deployed, ML artifacts are inherently dynamic as they learn from data, update their parameters, and participate in feedback loops with the environments they model. This dynamism calls for design frameworks that account for ongoing change, frameworks where interpretability enables continuous human oversight, cost-awareness ensures sustainable operation across repeated cycles, and value-sensitivity allows realignment as contexts evolve.

Limitations and future research

The dissertation has limitations that define concrete opportunities for future work. First, the digital maturity studies rely primarily on survey-based measurement. Future research could strengthen validity through longitudinal designs and complementary data sources, including ethically and institutionally appropriate digital trace measures where feasible. Second, while SHAP provides useful feature-attribution summaries, interpretability does not automatically guarantee appropriate human understanding or action. Future research should examine the "last mile" of interpretability like how explanations are communicated, understood, and used by non-technical stakeholders, such as teachers or policymakers, in real decision environments. Third, the governance analysis is situated in the EU context. Comparative research across jurisdictions could test the portability of the derived meta-requirements and clarify how institutional differences condition the feasibility of data democratization mechanisms.

This dissertation advances a design-oriented account of ML in IS. ML can strengthen IS research and practice by enabling operational measurement under constraint, revealing heterogeneity that matters for intervention and governance, and supporting systematic analysis of policy texts and stakeholder priorities. These benefits are most defensible when ML systems are treated as socio-technical decision artifacts, interpretable enough to justify decisions, feasible under real constraints on measurement and implementation, and aligned with the values embedded in organizational and institutional arrangements (Padmanabhan et al., 2022; Selbst et al., 2019). The three studies demonstrate that socio-technical ML design is not domain-specific but generalizable across IS contexts. The same principles of embedding interpretability for stakeholder comprehension, respecting costs for sustainable deployment, encoding values for ethical alignment, apply whether the context is educational assessment or regulatory policy. This suggests that "Socio-Technical ML Design" can function as a mid-range design theory applicable across IS domains where algorithmic systems mediate human decisions. The dissertation thus contributes not only substantive findings about digital maturity and data governance, but also a transferable design orientation for responsible ML in information systems.

Conclusión

Esta tesis doctoral se propuso abordar una tensión práctica y teórica en la investigación en Sistemas de Información (SI). El aprendizaje automático (Machine Learning, ML) ofrece capacidades potentes para la predicción y el descubrimiento de patrones; sin embargo, los sistemas de ML a menudo encuentran barreras para un uso legítimo y eficaz en entornos socio-técnicos, ya que pueden ser opacos, exigir elevados requisitos de datos e implementación, y estar desalineados con los valores institucionales. A través de tres estudios, la tesis sostiene que estos retos se comprenden mejor como problemas de diseño socio-técnico, más que como limitaciones puramente técnicas. La tesis central es que el ML en SI debe tratarse como infraestructura socio-técnica de decisión: un artefacto que participa en procesos de decisión y que, por tanto, debe evaluarse con criterios que van más allá de la precisión, incorporando interpretabilidad para los actores relevantes, viabilidad bajo restricciones reales y sensibilidad a valores (Padmanabhan et al., 2022; Selbst et al., 2019).

Los tres capítulos abordan dominios y métodos distintos, pero están unificados por una lógica de diseño común, en la medida en que cada uno construye un artefacto basado en ML orientado a apoyar la toma de decisiones bajo una restricción socio-técnica específica.

El Capítulo 1 se centra en la *viabilidad operativa bajo restricciones de medición*. Muestra que la adquisición de información puede tratarse como un problema de asignación de recursos y que el aprendizaje por refuerzo puede utilizarse para reducir la carga de medición preservando la calidad de las decisiones (Janisch et al., 2020; Sutton and Barto, 2018). El agente de DRL aprende qué ítems del cuestionario aportan mayor valor discriminativo, permitiendo a los profesionales obtener clasificaciones fiables de madurez sin administrar evaluaciones extensas, un requisito crítico para su despliegue en centros educativos y servicios para jóvenes.

El Capítulo 2 se centra en la *rendición de cuentas epistémica en investigación habilitada por ML*. Al combinar descubrimiento mediante ML interpretable con contrastación estadística confirmatoria, demuestra un flujo de trabajo disciplinado que aprovecha la flexibilidad del ML sin reducir la inferencia a mera minería de patrones (Shmueli, 2010; Lundberg and Lee, 2017). El enfoque de descubrimiento y confirmación garantiza que las relaciones identificadas mediante SHAP no sean solo correlaciones descriptivas, sino que puedan resistir el escrutinio inferencial, reforzando su utilidad para la teoría del desarrollo y el diseño de intervenciones.

El Capítulo 3 se centra en la *complejidad normativa en la gobernanza*. Al mapear los énfasis de valores de los distintos actores y vincularlos con mecanismos regulatorios implementables, muestra cómo el ML puede apoyar un análisis sensible a valores del diseño tecno-regulatorio (Friedman et al., 2006, 2019; European Union, 2023). El análisis basado en PLN hace explícitas tensiones que, de otro modo, podrían permanecer implícitas en el discurso de políticas públicas y las traduce en meta-requisitos que pueden orientar la implementación regulatoria.

En conjunto, estos estudios sustentan una síntesis de nivel superior: un ML responsable en SI requiere alinear el diseño del modelo con las condiciones institucionales de uso. El artefacto no es solo el modelo predictivo; es el arreglo socio-técnico completo que incluye qué se mide, cómo se generan las salidas, cómo se explican y cómo se traducen en acción bajo restricciones de gobernanza. De estos estudios se derivan tres contribuciones integradoras:

La interpretabilidad como puente entre salidas y acción: una primera contribución consiste en tratar la interpretabilidad como un puente funcional entre las salidas del modelo y la acción institucional. En los estudios de madurez digital, la interpretabilidad apoya un razonamiento relevante para la intervención al aclarar qué competencias impulsan las clasificaciones y cómo cambian las relaciones a lo largo de etapas del desarrollo (Doshi-Velez and Kim, 2017; Lundberg and Lee, 2017). En el estudio de gobernanza, la interpretabilidad permite la trazabilidad desde patrones computacionales en el texto hacia mecanismos concretos de política pública, posibilitando que las afirmaciones de valor se examinen como compromisos de diseño y no se acepten como resultados opacos. Esto es coherente con la idea de que las explicaciones deben evaluarse en función del propósito y la audiencia, y con las advertencias de que la explicación post hoc, por sí sola, no garantiza la rendición de cuentas (Rudin, 2019).

La viabilidad como parte del rendimiento: una segunda contribución consiste en conceptualizar la viabilidad como parte del rendimiento, y no como una preocupación de implementación posterior. Muchos entornos organizativos no pueden sostener prácticas de recopilación de datos que son costosas en tiempo, atención o carga de cumplimiento. El Capítulo 1 muestra que diseñar artefactos de apoyo a la decisión sensibles al coste puede marcar la diferencia entre un enfoque de ML teóricamente llamativo y uno prácticamente desplegable (Janisch et al., 2020; Sutton and Barto, 2018). En términos más amplios, la tesis sugiere que los artefactos socio-técnicos de ML deben evaluarse bajo restricciones realistas de medición y capacidad institucional, especialmente en ámbitos como educación, servicios públicos o sanidad, donde los recursos y la legitimidad están acotados.

Los valores como variables de diseño vinculadas a mecanismos: una tercera contribución es mostrar que los valores pueden tratarse como variables de diseño y no como comentario externo. Los valores entran en los sistemas de ML a través de decisiones sobre objetivos, proxies, umbrales y arreglos de gobernanza (Selbst et al., 2019; Mitchell et al., 2021). El Capítulo 3 extiende esta lógica al ámbito de políticas públicas al vincular los énfasis de valores de los actores con mecanismos de gobernanza implementables y al derivar meta-requisitos que traducen objetivos de democratización de datos en compromisos institucionales diseñables (Friedman et al., 2006, 2019; European Union, 2023). Esto ofrece un enfoque estructurado para conectar el análisis normativo con mecanismos socio-técnicos concretos en investigación SI.

Artefactos de ML y retroalimentación socio-técnica

Una característica crucial de los sistemas de ML en contextos de SI es su participación en bucles de retroalimentación. Las predicciones generadas por el ML no solo describen la realidad; intervienen en ella. Una evaluación de madurez digital, una vez entregada a un docente, condiciona las intervenciones que recibe el estudiante, lo que a su vez afecta a su comportamiento digital futuro y a evaluaciones posteriores. De forma similar, un marco de política diseñado mediante análisis de PLN sensible a valores generará respuestas de los actores que reconfiguran la siguiente iteración regulatoria. Esta reflexividad, en la que las salidas del ML se convierten en entradas de los sistemas sociales que modelan, sitúa al ML como infraestructura socio-técnica más que como una herramienta pasiva de medición. Reconocer esta dinámica de retroalimentación tiene varias implicaciones. En primer lugar, sugiere que los artefactos de ML requieren una gobernanza continua, en lugar de una validación única. En segundo lugar, implica que los principios de diseño socio-técnico deben incorporarse desde el inicio y no añadirse a posteriori tras el despliegue. En tercer lugar, motiva una agenda de investigación prospectiva que estudie cómo se comportan los artefactos de ML tras su despliegue, cómo se adaptan los actores a ellos y cómo coevolucionan los controles técnicos e institucionales con el tiempo (Faraj et al., 2018; Berente et al., 2021). A diferencia de los artefactos tradicionales de TI, que permanecen relativamente estables una vez desplegados, los artefactos de ML son intrínsecamente dinámicos: aprenden de los datos, actualizan sus parámetros y participan en bucles de retroalimentación con los entornos que modelan. Esta dinámica exige marcos de diseño que contemplen el cambio continuo, marcos en los que la interpretabilidad permita una supervisión humana sostenida, la sensibilidad al coste asegure una operación sostenible a lo largo de ciclos repetidos y la sensibilidad a valores posibilite realineaciones conforme evolucionan los contextos.

Limitaciones y futuras líneas de investigación

La tesis presenta limitaciones que abren oportunidades concretas para investigaciones futuras. En primer lugar, los estudios de madurez digital se apoyan principalmente en medición basada en encuestas. Trabajos futuros podrían reforzar la validez mediante diseños longitudinales y fuentes de datos complementarias, incluyendo, cuando sea viable, medidas de trazas digitales que sean éticamente aceptables y apropiadas desde el punto de vista institucional. En segundo lugar, aunque SHAP aporta resúmenes útiles de atribución de características, la interpretabilidad no garantiza automáticamente una comprensión humana adecuada ni una acción apropiada. Investigaciones futuras deberían examinar la "última milla" de la interpretabilidad: cómo se comunican, se comprenden y se utilizan las explicaciones por parte de actores no técnicos, como docentes o responsables de políticas, en entornos reales de decisión. En tercer lugar, el análisis de gobernanza se sitúa en el contexto de la Unión Europea. La investigación comparativa entre jurisdicciones podría poner a prueba la portabilidad de los meta-requisitos derivados y aclarar cómo las diferencias institucionales condicionan la viabilidad de los mecanismos de democratización de datos.

En conjunto, esta tesis avanza una lectura orientada al diseño del ML en SI. El ML puede fortalecer la investigación y la práctica en SI al permitir medición operativa bajo restricciones, revelar heterogeneidad relevante para la intervención y la gobernanza, y apoyar el análisis sistemático de textos de política pública y de prioridades de los actores. Estos beneficios son más defendibles cuando los sistemas de ML se tratan como artefactos socio-técnicos de decisión: lo suficientemente interpretables como para justificar decisiones, viables bajo restricciones reales de medición e implementación, y alineados con los valores incrustados en los arreglos organizativos e institucionales (Padmanabhan et al., 2022; Selbst et al., 2019). Los tres estudios muestran que el diseño socio-técnico de ML no es específico de un dominio, sino generalizable a distintos contextos de SI. Los mismos principios (incorporar interpretabilidad para la comprensión de los actores, respetar costes para un despliegue sostenible y codificar valores para una alineación ética) se aplican tanto si el contexto es la evaluación educativa como si lo es la política regulatoria. Esto sugiere que el "Diseño socio-técnico de ML" puede funcionar como una teoría de diseño de alcance medio aplicable en dominios de SI donde los sistemas algorítmicos median decisiones humanas. De este modo, la tesis contribuye no solo con hallazgos sustantivos sobre madurez digital y gobernanza de datos, sino también con una orientación de diseño transferible para un ML responsable en sistemas de información.

References

Abraham, R., Schneider, J., and Vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International journal of information management*, 49:424–438.

Akhlaghpour, S., Wu, J., Lapointe, L., and Pinsonneault, A. (2013). The ongoing quest for the IT artifact: Looking back, moving forward. *Journal of Information Technology*, 28(2):150–166.

Allen, M. (2017). Factor analysis: Varimax rotation. *The SAGE Encyclopaedia of Communication Research Methods*, 10(9781483381411):n191.

Allen, M. S., Walter, E. E., and Swann, C. (2019). Sedentary behaviour and risk of anxiety: a systematic review and meta-analysis. *Journal of Affective Disorders*, 242:5–13.

Ananny, M. and Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3):973–989.

Anderson, M., Jiang, J., et al. (2018). Teens, social media & technology 2018. *Pew research center*, 31(2018):1673–1689.

Andreassen, C. S., Pallesen, S., and Griffiths, M. D. (2017). The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. *Addictive behaviors*, 64:287–293.

Apley, D. W. and Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(4):1059–1086.

Arenas, A. E. and Yazdi, P. (2022). Towards a model of technology usage and digital maturity in children: a grounded-theory approach. In *ECIS*.

Arend, M. G. and Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on monte carlo simulation. *Psychological methods*, 24(1):1.

Arpaci, I. (2023). Predicting problematic smartphone use based on early maladaptive schemas by using machine learning classification algorithms. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, 41(3):634–643.

Baker, R. S., Martin, T., and Rossi, L. M. (2016). Educational data mining and learning analytics. *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications*, pages 379–396.

Bakola, L. N., Rizos, N. D., and Drigas, A. S. (2019). Icts for emotional and social skills development for children with adhd and asd co-existence. *International Journal of Emerging Technologies in Learning*, 14(5).

Baldwin, C. Y. and Clark, K. B. (2000). *Design Rules, Volume 1: The Power of Modularity*. MIT Press, Cambridge, MA.

Barocas, S. and Selbst, A. D. (2016). Big data's disparate impact. *California law review*, pages 671–732.

Bélanger, F. and Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly*, 35(4):1017–1042.

Belsky, J. and Pluess, M. (2009). Beyond diathesis stress: differential susceptibility to environmental influences. *Psychological bulletin*, 135(6):885.

Benbasat, I. and Zmud, R. W. (2003). The identity crisis within the IS discipline: Defining and communicating the discipline's core properties. *MIS Quarterly*, 27(2):183–194.

Benbya, H., Nan, N., Tanriverdi, H., and Yoo, Y. (2020). Complexity and information systems research in the emerging digital world. *MIS quarterly*, 44(1):1–17.

Berente, N., Gu, B., Recker, J., and Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, 45(3):1433–1450.

Beyens, I., Pouwels, J. L., van Driel, I. I., Keijsers, L., and Valkenburg, P. M. (2020). The effect of social media on well-being differs from adolescent to adolescent. *Scientific reports*, 10(1):10763.

Blakemore, S.-J. and Mills, K. L. (2014). Is adolescence a sensitive period for sociocultural processing? *Annual review of psychology*, 65(1):187–207.

Boers, E., Afzali, M. H., Newton, N., and Conrod, P. (2019). Association of screen time and depression in adolescence. *JAMA pediatrics*, 173(9):853–859.

Bovens, M. (2007). Analysing and assessing public accountability: A conceptual framework. *European Law Journal*, 13(4):447–468.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

Burrow, A. L. and Rainone, N. (2017). How many likes did I get?: Purpose moderates links between positive social media feedback and self-esteem. *Journal of Experimental Social Psychology*, 69:232–236.

Burton-Jones, A. and Straub Jr, D. W. (2006). Reconceptualizing system usage: An approach and empirical test. *Information systems research*, 17(3):228–246.

California Consumer Privacy Act (2023). California consumer privacy act (as amended by the CPRA). `https://www.oag.ca.gov/privacy/ccpa`.

Campion, S. (2023). Europeanparliment/eubert.

Casey, B. J., Getz, S., and Galvan, A. (2008). The adolescent brain. *Developmental review*, 28(1):62–77.

Cecez-Kecmanovic, D. (2019). The resistible rise of the digital surveillance economy: A call for action. *Journal of Information Technology*, 34(1):81–83.

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., and Androutsopoulos, I. (2019). Large-scale multi-label text classification on eu legislation. *arXiv preprint arXiv:1906.02192*.

Chaplin, T. M. and Aldao, A. (2013). Gender differences in emotion expression in children: a meta-analytic review. *Psychological bulletin*, 139(4):735.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.

Christensen, R. N., Siddiqui, A., Valogianni, K., Florack, A., and Hubert, M. (2025). Digital maturity of young people–unfolding dependencies between core dimensions and national context as external factor by using a machine-learning approach. *Computers in Human Behavior Reports*, page 100778.

Clifford Chance (2023). Entry into force of the eu data act: What are the key requirements?

Cobbe, J., Veale, M., and Singh, J. (2023). Understanding accountability in algorithmic supply chains. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pages 1186–1197.

Crone, E. A. and Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nature reviews neuroscience*, 13(9):636–650.

De Asis, K., Hernandez-Garcia, J., Holland, G., and Sutton, R. (2018). Multi-step reinforcement learning: A unifying algorithm. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

de Reuver, M., Sørensen, C., and Basole, R. C. (2018). The digital platform: A research agenda. *Journal of Information Technology*, 33(2):124–135.

Department of Health and Human Services (2020). 21st century cures act: Interoperability, information blocking, and the onc health it certification program final rule. `https://www.federalregister.gov/documents/2020/05/01/2020-07419/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification`.

Dhir, A., Chen, S., and Nieminen, M. (2016). Psychometric validation of the compulsive internet use scale: relationship with adolescents' demographics, ict accessibility, and problematic ict use. *Social Science Computer Review*, 34(2):197–214.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer.

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1):114–126.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1):27–46.

Doshi-Velez, F. and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Dutton, W. H. (2014). Putting things to work: Social and policy challenges for the internet of things. *info*, 16(3):1–21.

EU Data Act Chapter 36 (2024). Article 36 data act: Essential requirements for smart contracts.

EU Data Act Chapter 5 (2024). Data act chapter v: Making data available to public sector bodies in situations of exceptional need.

European Commission (2018). Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions "Towards a common European data space"'.

European Commission (2020). A european strategy for data.

European Commission (2021). Data Act amended rules on the legal protection of databases.

European Commission (2022a). A European Strategy for data.

European Commission (2022b). Regulation of the European parliament and of the council on harmonised rules on fair access to and use of data (Data Act).

European Commission (2025). Commission publishes guidelines on the protection of minors. `https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-protection-minors`. Accessed: 2026-01-08.

European Parliament and Council of the European Union (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union L 119/1.

European Union (2022a). Regulation (eu) 2022/2065 of the european parliament and of the council of 19 october 2022 on a single market for digital services and amending directive 2000/31/ec (digital services act).

European Union (2022b). Regulation (eu) 2022/868 on european data governance (data governance act). Official Journal of the European Union.

European Union (2023). Regulation (eu) 2023/2854 on fair access to and use of data (data act). Official Journal of the European Union.

Eurostat (2019). Statistics explained.

Faraj, S. and Azad, B. (2012). The materiality of technology: An affordance perspective. *Materiality and organizing: Social interaction in a technological world*, 237(1):237–258.

Faraj, S., Pachidi, S., and Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1):62–70.

Fardouly, J., Diedrichs, P. C., Vartanian, L. R., and Halliwell, E. (2015). Social comparisons on social media: The impact of facebook on young women's body image concerns and mood. *Body image*, 13:38–45.

Farrell, J. and Klemperer, P. (2007). Coordination and lock-in: Competition with switching costs and network effects. *Handbook of industrial organization*, 3:1967–2072.

Feng, Q. A. and Shanthikumar, J. G. (2023). The framework of parametric and nonparametric operational data analytics. *Production and Operations Management*, 32(9):2685–2703.

Fernández-Loría, C., Provost, F., and Han, X. (2022). Explaining data-driven decisions made by ai systems: The counterfactual approach. *MIS Quarterly*, 46(3):1635–1660.

Fox, J. (2007). The uncertain relationship between transparency and accountability. *Development in Practice*, 17(4-5):663–671.

Fraillon, J. (2024). An international perspective on digital literacy: Results from icils 2023. *International Association for the Evaluation of Educational Achievement*.

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., and Duckworth, D. (2020). *Preparing for life in a digital world: IEA international computer and information literacy study 2018 international report*. Springer Nature.

Frauenberger, C., Good, J., and Keay-Bright, W. (2011). Designing technology for children with special needs: bridging perspectives through participatory design. *CoDesign*, 7(1):1–28.

Friedman, B., Hendry, D. G., and Borning, A. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. The MIT Press.

Friedman, B., Kahn, P. H., and Borning, A. (2014). Value sensitive design and information systems. *Early engagement and new technologies: Opening up the laboratory*, 16.

Friedman, B., Kahn, P. H., Borning, A., and Huldtgren, A. (2013). Value sensitive design and information systems. *Early engagement and new technologies: Opening up the laboratory*, pages 55–95.

Friedman, B., Kahn, P. H., and Hagman, J. (2008). Value sensitive design and information systems. *The Handbook of Information and Computer Ethics*, pages 69–101.

Friedman, B., Kahn, Jr., P. H., and Borning, A. (2006). Value sensitive design and information systems. In Zhang, P. and Galletta, D., editors, *Human-Computer Interaction in Management Information Systems: Foundations*, pages 348–372. M. E. Sharpe, Armonk, NY.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Gargett, A., Firth, R., and Aletras, N. (2020). Legalops: a summarization corpus of legal opinions. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 2117–2120. IEEE.

Gazzaley, A. and Rosen, L. D. (2016). *The distracted mind: Ancient brains in a high-tech world*. Mit Press.

GDPR Article 5 (2016). Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 (general data protection regulation), article 5: Principles relating to processing of personal data. Official Journal of the European Union L119.

Ghazawneh, A. and Henfridsson, O. (2013). Balancing platform control and external contribution in third-party development: The boundary resources model. *MIS Quarterly*, 37(2):655–676.

Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings.

Gnambs, T. (2021). The development of gender differences in information and communication technology (ict) literacy in middle adolescence. *Computers in Human Behavior*, 114:106533.

Goldstein, A., Kapelner, A., Bleich, J., and Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *journal of Computational and Graphical Statistics*, 24(1):44–65.

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT press Cambridge.

Goodhue, D. L. and Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS quarterly*, pages 213–236.

Greenberger, E., Josselson, R., Knerr, C., and Knerr, B. (1975). The measurement and structure of psychosocial maturity. *Journal of Youth and Adolescence*, 4(2):127–143.

Greenberger, E. and Sørensen, A. B. (1974). Toward a concept of psychosocial maturity. *Journal of Youth and Adolescence*, 3(4):329–358.

Gregor, S. and Benbasat, I. (1999). Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Quarterly*, 23(4):497–530.

Gregory, R. W., Kaganer, E., Henfridsson, O., and Ruch, T. J. (2018). IT Consumerization and the Transformation of IT Governance. *MIS Quarterly*, 42(4):1225–9.

Griva, A., Kotsopoulos, D., Karagiannaki, A., and Zamani, E. D. (2023). What do growing early-stage digital start-ups look like? a mixed-methods approach. *International Journal of Information Management*, 69:102427.

GSMA (2025). The mobile economy 2024.

Guggenberger, T. M., Schlueter Langdon, C., and Otto, B. (2025). Data spaces as meta-organisations. *European Journal of Information Systems*, pages 1–21.

Han, X., Lu, S., Wang, X., and Cui, N. (2025). Effectiveness of online education during the covid-19 pandemic: Evidence from chinese universities. *Production and Operations Management*, page 10591478251361979.

Hanseth, O. and Lyytinen, K. (2010). Design theory for dynamic complexity in information infrastructures: The case of building internet. *Journal of Information Technology*, 25(1):1–19.

Hargittai, E. (2001). Second-level digital divide: Mapping differences in people's online skills. *arXiv preprint cs/0109068*.

Hasebrink, U., Livingstone, S., Haddon, L., and Olafsson, K. (2009). *Comparing children's online opportunities and risks across Europe: Cross-national comparisons for EU Kids Online*. EU Kids Online, The London School of Economics and Political Science.

Hastie, T. J. (2017). Generalized additive models. *Statistical models in S*, pages 249–307.

Helsper, E., Kalmus, V., Hasebrink, U., Sagvari, B., and De Haan, J. (2013). *Country classification: Opportunities, risks, harm and parental mediation*. EU Kids Online, LSE.

Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design science in Information Systems Research. *MIS Quarterly*, 28(1):75–105.

Hill, D., Ameenuddin, N., Reid Chassiakos, Y. L., Cross, C., Hutchinson, J., Levine, A., Boyd, R., Mendelson, R., Moreno, M., Swanson, W. S., et al. (2016). Media and young minds. *Pediatrics*, 138(5).

Hinduja, S. and Patchin, J. W. (2010). Bullying, cyberbullying, and suicide. *Archives of suicide research*, 14(3):206–221.

Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Hockenberry, M. J., Hooke, M. C., McCarthy, K., and Gregurich, M. A. (2011). Sickness behavior clustering in children with cancer. *Journal of Pediatric Oncology Nursing*, 28(5):263–272.

Hodapp, D. and Hanelt, A. (2022). Interoperability in the era of digital innovation: An information systems research agenda. *Journal of Information Technology*, 37(4):407–427.

Hofmans, L., van der Stappen, A., and van den Bos, W. (2024). Developmental structure of digital maturity. *Computers in Human Behavior*, 157:108239.

Hu, J. and Yu, R. (2021). The effects of ict-based social media on adolescents' digital reading performance: A longitudinal study of pisa 2009, pisa 2012, pisa 2015 and pisa 2018. *Computers & Education*, 175:104342.

Hyde, J. S. (2005). The gender similarities hypothesis. *American psychologist*, 60(6):581.

Jackson, L. A., Fitzgerald, H. E., Zhao, Y., Kolenic, A., Von Eye, A., and Harold, R. (2008). Information technology use and children's psychological well-being. *CyberPsychology & Behavior*, 11(6):755–757.

Janisch, J., Pevnỳ, T., and Lisỳ, V. (2020). Classification with costly features as a sequential decision-making problem. *Machine Learning*, 109(8):1587–1615.

Janssen, M., Charalabidis, Y., and Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information systems management*, 29(4):258–268.

Jones, L. M. and Mitchell, K. J. (2016). Defining and measuring youth digital citizenship. *New Media & Society*, 18(9):2063–2079.

Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260.

Katz, D. M., Hartung, D., Gerlach, L., Jana, A., and Bommarito II, M. J. (2023). Natural language processing in the legal domain. *arXiv preprint arXiv:2302.12039*.

Kerckaert, S., Vanderlinde, R., and van Braak, J. (2015). The role of ict in early childhood education: Scale development and research on ict use and influencing factors. *European Early Childhood Education Research Journal*, 23(2):183–199.

Khatri, V. and Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1):148–152.

Kim, J.-Y. (2025). Data portability and interoperability between digital platforms. *Journal of Economics & Management Strategy*.

Kingma, D. P. and Ba, J. (2014). Adam:a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Koch, T., Laaber, F., and Florack, A. (2024). Socioeconomic status and young people's digital maturity: The role of parental mediation. *Computers in Human Behavior*, 154:108157.

Kornilova, A. and Eidelman, V. (2019). Billsum: A corpus for automatic summarization of us legislation. *arXiv preprint arXiv:1910.00523*.

Kronblad, C., Essén, A., and Mähring, M. (2024). When justice is blind to algorithms: Multilayered blackboxing of algorithmic decision-making in the public sector. *MIS Quarterly*, 48(4):1637–1662.

Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D. S., Lin, N., Shablack, H., Jonides, J., and Ybarra, O. (2013). Facebook use predicts declines in subjective well-being in young adults. *PloS one*, 8(8):e69841.

Laaber, F., Florack, A., Koch, T., and Hubert, M. (2023). Digital maturity: Development and Validation of the Digital Maturity Inventory (DIMI). *Computers in Human Behavior*, 143:107709.

Laaber, F., Koch, T., Florack, A., and Hubert, M. (2025). Digital maturity: enabling a beneficial use of digital technologies by children and adolescents. In *Young People in Digital Environments*, pages 251–269. Edward Elgar Publishing.

Laaber, F., Koch, T., Hubert, M., and Florack, A. (2024). Young people's digital maturity relates to different forms of well-being through basic psychological need satisfaction and frustration. *Computers in Human Behavior*, 152:108077.

Latham and Watkins (2025). China clarifies privacy and data security requirements in network data security management regulations. `https://www.lw.com/admin/upload/SiteAttachments/China-Clarifies-Privacy-and-Data-Security-Requirements-in-Network-Data-Security-Management-Regulations.pdf`.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Lefebvre, H., Legner, C., and Fadler, M. (2021). Data democratization: toward a deeper understanding. In *ICIS*.

Leonardi, P. M. (2011). When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly*, pages 147–167.

Leonardi, P. M. (2012). Materiality, sociomateriality, and socio-technical systems: What do these terms mean? how are they different? do we need them. *Materiality and organizing: Social interaction in a technological world*, 25(10):1093.

Li, Y., Thomas, M. A., and Liu, D. (2021). From semantics to pragmatics: where is can lead in natural language processing (nlp) research. *European Journal of Information Systems*, 30(5):569–590.

Liang, W., Yang, Y., Zhang, Y., Zhang, L., Wang, C., and Chen, J. (2023). Clique hierarchy moderates the association between social preference and defending behaviors in early adolescence: the role of gender differences. *Journal of Youth and Adolescence*, 52(11):2285–2299.

Lin, E., Chen, Q., and Qi, X. (2020). Deep reinforcement learning for imbalanced classification. *Applied Intelligence*, 50(8):2488–2502.

Lippi, M., Pałka, P., Contissa, G., Lagioia, F., Micklitz, H.-W., Sartor, G., and Torroni, P. (2019). Claudette: an automated detector of potentially unfair clauses in online terms of service. *Artificial Intelligence and Law*, 27:117–139.

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3):31–57.

Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation forest. pages 413–422. IEEE.

Livingstone, S. (2012). Critical reflections on the benefits of ICT in education. *Oxford Review of Education*, 38(1):9–24.

Livingstone, S., Haddon, L., Görzig, A., and Ólafsson, K. (2011). Risks and safety on the internet: the perspective of european children: full findings and policy implications from the eu kids online survey of 9-16 year olds and their parents in 25 countries.

Livingstone, S. and Helsper, E. (2007). Gradations in digital inclusion: Children, young people and the digital divide. *New media & society*, 9(4):671–696.

Livingstone, S. and Helsper, E. (2010). Balancing opportunities and risks in teenagers' use of the internet: The role of online skills and internet self-efficacy. *New media & society*, 12(2):309–329.

Livingstone, S., Mascheroni, G., and Staksrud, E. (2018). European research on children's internet use: Assessing the past and anticipating the future. *New Media & Society*, 20(3):1103–1122.

Livingstone, S., Mascheroni, G., and Stoilova, M. (2023). The outcomes of gaining digital skills for young people's lives and wellbeing: A systematic evidence review. *New media & society*, 25(5):1176–1202.

Logg, J. M., Minson, J. A., and Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151:90–103.

Long, J. S. and Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3):217–224.

Lukoff, K., Yu, C., Kientz, J., and Hiniker, A. (2018). What makes smartphone use meaningful or meaningless? *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):1–26.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1):56–67.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

Luo, J., Liang, L., and Li, H. (2020). The divergent roles of social media in adolescents' academic performance. *Journal of Research in Childhood Education*, 34(2):167–182.

Lyngs, U., Lukoff, K., Slovak, P., Binns, R., Slack, A., Inzlicht, M., Van Kleek, M., and Shadbolt, N. (2019). Self-control in cyberspace: Applying dual systems theory to a review of digital self-control tools. In *proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–18.

Lyytinen, K. and King, J. L. (2006). Editorial—standards making: A critical research frontier for information systems research. *MIS Quarterly*, 30:289–310.

Ma, C. M. (2022). Relationships between social networking sites use and self-esteem: the moderating role of gender. *International journal of environmental research and public health*, 19(18):11462.

Maeneja, R., Rato, J., and Ferreira, I. S. (2025). How is the digital age shaping young minds? a rapid systematic review of executive functions in children and adolescents with exposure to ict. *Children*, 12(5):555.

Majchrzak, A. and Markus, M. L. (2012). Technology affordances and constraints in management information systems (mis). *Encyclopedia of Management Theory,(Ed: E. Kessler), Sage Publications, Forthcoming*.

Malhotra, N. K., Kim, S. S., and Agarwal, J. (2004). Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model. *Information Systems Research*, 15(4):336–355.

Manning, C. (1999). *Foundations of statistical natural language processing*. MIT press.

Marchand, D. A., Kettinger, W. J., and Rollins, J. D. (2000). Information orientation: people, technology and the bottom line. *MIT Sloan Management Review*, 41(4):69.

Marcolin, B. L., Compeau, D. R., Munro, M. C., and Huff, S. L. (2000). Assessing user competence: Conceptualization and measurement. *Information systems research*, 11(1):37–60.

Markus, M. L. and Silver, M. S. (2008). A foundation for the study of it effects: A new look at desanctis and poole's concepts of structural features and spirit. *Journal of the Association for Information systems*, 9(10):5.

Markus, M. L., Steinfield, C. W., and Wigand, R. T. (2006a). Industry-wide information systems standardization as collective action: the case of the us residential mortgage industry. *MIS quarterly*, pages 439–465.

Markus, M. L., Steinfield, C. W., Wigand, R. T., and Minton, T. (2006b). Industry-wide information systems standardization as collective action: The case of the u.s. residential mortgage industry. *MIS Quarterly*, 30:439–465.

Marsh, J., Plowman, L., Yamada-Rice, D., Bishop, J., and Scott, F. (2020). Digital play: A new classification. In *Digital Play and Technologies in the Early Years*, pages 20–31. Routledge.

Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, 160(4):835–850.

Meng, X.-L. (2024). Data democratization: An ecosystemic contemplation and coordination. *Harvard Data Science Review*, (Special Issue 4).

Mikalef, P., Conboy, K., Lundström, J. E., and Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side'of AI.

Mingers, J. (2001). Combining is research methods: towards a pluralist methodology. *Information systems research*, 12(3):240–259.

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., and Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annual review of statistics and its application*, 8(1):141–163.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, New York.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.

Morris, M. G. and Venkatesh, V. (2000). Age differences in technology adoption decisions: Implications for a changing work force. *Personnel psychology*, 53(2):375–403.

Mullainathan, S. and Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

Nesi, J. and Prinstein, M. J. (2015). Using social media for social comparison and feedback-seeking: Gender and popularity moderate associations with depressive symptoms. *Journal of abnormal child psychology*, 43(8):1427–1438.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79(1):119–158.

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, Stanford, CA.

Nussbaum, M. C. (2011). *Creating Capabilities: The Human Development Approach*. Harvard University Press, Cambridge, MA.

Odgers, C. L. and Jensen, M. R. (2020). Annual research review: Adolescent mental health in the digital age: Facts, fears, and future directions. *Journal of Child Psychology and Psychiatry*, 61(3):336–348.

Orben, A. and Matias, J. N. (2025). Fixing the science of digital technology harms. *Science*, 388(6743):152–155.

Orben, A. and Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature human behaviour*, 3(2):173–182.

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization studies*, 28(9):1435–1448.

Orlikowski, W. J. and Iacono, C. S. (2001). Desperately seeking the IT artifact in IS research: A call to theorizing the IT artifact. *MIS Quarterly*, 25(2):121–134.

Padmanabhan, B., Sahoo, N., Burton-Jones, A., et al. (2022). Machine learning in Information Systems Research. *MIS Quarterly*, 46(1):iii–xix.

Pandya, A. and Lodha, P. (2021). Social connectedness, excessive screen time during covid-19 and mental health: a review of current evidence. *Frontiers in Human Dynamics*, 3:45.

Patchin, J. W. and Hinduja, S. (2017). Digital self-harm among adolescents. *Journal of Adolescent Health*, 61(6):761–766.

Pavlou, P. A. and Gefen, D. (2004). Building effective online marketplaces with institution-based trust. *Information Systems Research*, 15(1):37–59.

Polyviou, A., Pouloudi, N., Pramatari, K., and Silva, L. O. (2024). A dna helix analogy for interdependent mixed-methods research: Enabling cross-fertilizations and interim meta-inferences. Forthcoming.

Ponti, M. (2023). Screen time and preschool children: Promoting health and development in a digital world. *Paediatrics & child health*, 28(3):184–192.

Prensky, M. (2001). Digital natives, digital immigrants part 2: Do they really think differently? *On the horizon*, 9(6):1–6.

Przybylski, A. K. and Weinstein, N. (2017). A large-scale test of the goldilocks hypothesis: quantifying the relations between digital-screen use and the mental well-being of adolescents. *Psychological science*, 28(2):204–215.

Punamäki, R.-L., Wallenius, M., Nygård, C.-H., Saarni, L., and Rimpelä, A. (2007). Use of information and communication technology (ict) and perceived health in adolescence: the role of sleeping habits and waking-time tiredness. *Journal of Adolescence*, 30(4):569–585.

Rabelo, J., Goebel, R., Kim, M.-Y., Kano, Y., Yoshioka, M., and Satoh, K. (2022). Overview and discussion of the competition on legal information extraction/entailment (coliee) 2021. *The Review of Socionetwork Strategies*, 16(1):111–133.

Radesky, J., Christakis, D., Hill, D., Ameenuddin, N., Chassiakos, Y., Cross, C., Radesky, J., Hutchinson, J., Boyd, R., Mendelson, R., et al. (2016). Media and young minds. *Pediatrics*, 138(5).

Rai, A. (2020). Explainable ai: From black box to glass box. *Journal of the academy of marketing science*, 48(1):137–141.

Reid Chassiakos, Y. L., Radesky, J., Christakis, D., Moreno, M. A., Cross, C., Hill, D., Ameenuddin, N., Hutchinson, J., Levine, A., Boyd, R., et al. (2016). Children and adolescents and digital media. *Pediatrics*, 138(5).

Reuters (2024a). China relaxes security review rules for some data exports. `https://www.reuters.com/technology/cybersecurity/chinas-cyberspace-regulator-issues-rules-facilitate-cross-border-data-flow\protect\discretionary{\char\hyphenchar\font}{}{}2024-03-22/`.

Reuters (2024b). China's revised and more stringent state secrets law takes effect. `https://www.reuters.com/legal/legalindustry/chinas-revised-more-stringent-state-secrets-law-takes-effect-2024-05-07/`.

Rideout, V. and Robb, M. B. (2020). The common sense census: Media use by kids age zero to eight, 2020.

Robert, L. P., Pierce, C., Marquis, L., Kim, S., and Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human-Computer Interaction*, 35(5-6):545–575.

Rose, A. J. and Rudolph, K. D. (2006). A review of sex differences in peer relationship processes: potential trade-offs for the emotional and behavioral development of girls and boys. *Psychological bulletin*, 132(1):98.

Rose, T., Barker, M., Jacob, C. M., Morrison, L., Lawrence, W., Strömmer, S., Vogel, C., Woods-Townsend, K., Farrell, D., Inskip, H., et al. (2017). A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. *Journal of Adolescent Health*, 61(6):669–677.

Rubinfeld, D. (2023). Data portability and interoperability: An e.u.–u.s. comparison. *European Business Law Review*, 34(5):877–904.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215.

Sarker, S., Chatterjee, S., Xiao, X., and Elbanna, A. (2019). The sociotechnical axis of cohesion for the is discipline: Its historical legacy and its continued relevance. *MIS Quarterly*, 43(3):695–720.

Schulze, L., Trenz, M., Cai, Z., and Tan, C.-W. (2023). Conducting online focus groups: Practical advice for information systems researchers. *Communications of the Association for Information Systems*, 52:385–428.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., and Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 59–68.

Sen, A. (1999). *Development as Freedom*. Oxford University Press, Oxford.

Shapiro, C. and Varian, H. R. (1999). *Information rules: A strategic guide to the network economy*. Harvard Business Press.

Sherazi, S. W. A., Bae, J.-W., and Lee, J. Y. (2021). A soft voting ensemble classifier for early prediction and diagnosis of occurrences of major adverse cardiovascular events for stemi and nstemi during 2-year follow-up in patients with acute coronary syndrome. *PloS one*, 16(6):e0249338.

Shmueli, G. (2010). To explain or to predict? *Statistical science*, pages 289–310.

Shmueli, G. and Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly*, pages 553–572.

Shreem, S. S., Turabieh, H., Al Azwari, S., and Baothman, F. (2022). Enhanced binary genetic algorithm as a feature selection to predict student performance. *Soft Computing*, 26(4):1811–1823.

Sidorova, A., Evangelopoulos, N., Valacich, J. S., and Ramakrishnan, T. (2008). Uncovering the intellectual core of the information systems discipline. *MIS Quarterly*, 32(3):467–482.

Silvers, J. A. (2022). Adolescence as a pivotal period for emotion regulation development. *Current opinion in psychology*, 44:258–263.

Skadden, Arps, Slate, Meagher & Flom LLP (2025). Eu data act: Three months to go before new rules on data access and use apply.

Sleimi, A. (2018). Automated extraction of semantic legal metadata using natural language processing. *2018 IEEE 26th International Requirements Engineering Conference (RE)*, pages 124–135.

Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S., and Hasebrink, U. (2020). Eu kids online 2020: Survey results from 19 countries.

Smith, H. J., Dinev, T., and Xu, H. (2011). Information privacy research: an interdisciplinary review. *MIS quarterly*, pages 989–1015.

Song, D., Gao, S., He, B., and Schilder, F. (2022a). On the effectiveness of pre-trained language models for legal natural language processing: An empirical study. *IEEE Access*, 10:75835–75858.

Song, D., Vold, A., Madan, K., and Schilder, F. (2022b). Multi-label legal document classification: A deep learning-based approach with label-attention and domain-specific pre-training. *Information Systems*, 106:101718.

Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in cognitive sciences*, 9(2):69–74.

Steinberg, L. and Cauffman, E. (1996). Maturity of judgment in adolescence: Psychosocial factors in adolescent decision making. *Law and Human Behavior*, 20(3):249–272.

Suttle, W., Zhang, K., Yang, Z., Liu, J., and Kraemer, D. (2021). Reinforcement learning for cost-aware markov decision processes. In *International Conference on Machine Learning*, pages 9989–9999. PMLR.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 2nd edition.

Sze, V., Chen, Y.-H., Yang, T.-J., and Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12):2295–2329.

Tan, J. and Gu, S. (2025). China clarifies cross-border data transfer rules: Practical guidance for compliance. `https://www.arnoldporter.com/en/perspectives/advisories/2025/06/china-clarifies-cross-border-data-transfer-rules`.

Tilson, D., Lyytinen, K., and Sørensen, C. (2010). Research commentary—digital infrastructures: The missing is research agenda. *Information systems research*, 21(4):748–759.

Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in human behavior*, 26(3):277–287.

Treem, J. W. and Leonardi, P. M. (2013). Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association. *Annals of the International Communication Association*, 36(1):143–189.

Tsetsi, E. and Rains, S. A. (2017). Smartphone internet access and use: Extending the digital divide and usage gap. *Mobile Media & Communication*, 5(3):239–255.

Turel, O. and Serenko, A. (2012). The benefits and dangers of enjoyment with social networking websites. *European Journal of Information Systems*, 21(5):512–528.

Twenge, J. M. and Campbell, W. K. (2018). Associations between screen time and lower psychological well-being among children and adolescents: Evidence from a population-based study. *Preventive medicine reports*, 12:271–283.

Twenge, J. M. and Martin, G. N. (2020). Gender differences in associations between digital media use and psychological well-being: Evidence from three large datasets. *Journal of adolescence*, 79:91–102.

Vaast, E., Safadi, H., Lapointe, L., and Negoita, B. (2017). Social media affordances for connective action. *MIS quarterly*, 41(4):1179–1206.

Valkenburg, P. M. and Peter, J. (2013). The differential susceptibility to media effects model. *Journal of communication*, 63(2):221–243.

Van Deursen, A. and Van Dijk, J. (2011). Internet skills and the digital divide. *New media & society*, 13(6):893–911.

Van Deursen, A. J. and Helsper, E. J. (2015). The third-level digital divide: Who benefits most from being online? In *Communication and information technologies annual*, volume 10, pages 29–52. Emerald Group Publishing Limited.

Van Deursen, A. J. and Van Dijk, J. A. (2014). The digital divide shifts to differences in usage. *New media & society*, 16(3):507–526.

Van Dijk, J. (2005). The deepening divide, inequality in the information society. sage publications.

Van Dijk, J. (2020). *The digital divide*. John Wiley & Sons.

Vandoninck, S., d'Haenens, L., and Roe, K. (2013). Online risks: Coping strategies of less resilient children and teenagers across europe. *Journal of children and media*, 7(1):60–78.

Varma, S. and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, 7(1):91.

Venkatesh, V., Brown, S. A., and Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS quarterly*, pages 21–54.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478.

Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., and Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? a critical review. *Social Issues and Policy Review*, 11(1):274–302.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272.

Vissenberg, J., d'Haenens, L., and Livingstone, S. (2022). Digital literacy and online resilience as facilitators of young people's well-being? *European Psychologist*.

Vogels, E. A., Gelles-Watnick, R., and Massarat, N. (2022). Teens, social media and technology 2022. *Pew Research Center*, 10.

Volkoff, O. and Strong, D. M. (2013). Critical realism and affordances: Theorizing it-associated organizational change processes. *MIS quarterly*, pages 819–834.

Wachter, S. Mittelstadt, B. and Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, 7(2):76–99.

Wang, R. Y. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4):5–33.

Wang, T., Jin, F., Hu, Y. J., Feng, L., and Cheng, Y. (2024). Making early and accurate deep learning predictions to help disadvantaged individuals in medical crowdfunding. *Production and Operations Management*.

Weber, T. A. (2014). Intermediation in a sharing economy: insurance, moral hazard, and rent extraction. *Journal of Management Information Systems*, 31(3):35–71.

Williams, A. (2010). *The ethos of Europe: Values, law and justice in the EU*. Cambridge University Press.

Williams, R., Cloete, R., Cobbe, J., et al. (2022). From transparency to accountability of intelligent systems: Moving beyond aspirations. *Data & Policy*, 4:e14.

Wilmer Hale (2023). Switching and interoperability requirements for cloud and other data processing services under the eu data act.

Zhang, J. and Chen, L. (2019). Clustering-based undersampling with random over sampling examples and support vector machine for imbalanced classification of breast cancer diagnosis. *Computer Assisted Surgery*, 24(sup2):62–72.

Zhang, X. (2017). Knowledge management system use and job performance. *MIS quarterly*, 41(3):811–A5.

Zhang, Z., Shi, L., and Zhou, D.-X. (2024). Classification with deep neural networks and logistic loss. *Journal of Machine Learning Research*, 25(125):1–117.

Zheng, L., Guha, N., Anderson, B. R., Henderson, P., and Ho, D. E. (2021). When does pretraining help? assessing self-supervised learning for law and the casehold dataset of 53,000+ legal holdings. In *Proceedings of the eighteenth international conference on artificial intelligence and law*, pages 159–168.

Zhu, Y., Mottaghi, R., Kolve, E., Lim, J. J., Gupta, A., Fei-Fei, L., and Farhadi, A. (2017). Target-driven visual navigation in indoor scenes using deep reinforcement learning. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 3357–3364. IEEE.

Zhu, Z., Lin, K., Jain, A. K., and Zhou, J. (2023). Transfer learning in deep reinforcement learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11):13344–13362.

Zimmermann, P. and Iwanski, A. (2014). Emotion regulation from early adolescence to emerging adulthood and middle adulthood: Age differences, gender differences, and emotion-specific developmental variations. *International journal of behavioral development*, 38(2):182–194.

Zou, Q., Xie, S., Lin, Z., Wu, M., and Ju, Y. (2016). Finding the best classification threshold in imbalanced classification. *Big Data Research*, 5:2–8.

Appendix

A.1 Appendix 1

A.1.1 Data Provenance, Governance, and Access

Origin and Stewardship. The data analyzed in this study originate from the DIGYMATEX project, which helped establish the Digital Maturity Inventory (DIMI) for adolescents in multiple European countries. The authors were affiliated with the project but did not conduct primary data collection at the sites. We received a de-identified research extract under a data-sharing agreement that specified scope of use and security controls. The DIMI instrument and its validation are described in Laaber et al. (2023).

Sampling and Administration. Participating institutions recruited adolescents aged 11–18 through partner schools and youth organizations. Surveys were administered online via a secure platform, in the local language, following site-specific protocols approved by each institution's ethics board. Country samples and descriptive statistics are reported in the methods section. No direct identifiers (names, email addresses, IP addresses) were included in the research extract.

Instrument and Translation. The survey captured DIMI dimensions covering digital literacy, risk awareness, individual growth, respect towards others, digital citizenship behavior, support-seeking, autonomy of choice, autonomy within digital context, regulation of aggressive impulses, and regulation of negative emotions. Country teams used standardized translations; where applicable, translations and back-translations were reviewed by native speakers with domain expertise to maintain conceptual equivalence with the source items.

Ethics and Consent. All sites obtained prior ethical approval and implemented informed consent procedures consistent with national regulations for research with minors. The appendix materials include (i) parental informed consent forms (German/English), (ii) adolescent assent forms (German/English), and (iii) institutional ethics approvals and positive votes from the relevant committees (e.g., University of Vienna ethics oversight). These documents specify eligibility, voluntary participation, the right to withdraw, and contacts for questions or complaints.

De-identification and Security. Prior to release, site teams removed direct identifiers, applied pseudonymous study IDs, and stored linkage files separately at the local institution. The shared dataset contained only derived or survey variables needed for analysis. Access was limited to the study team via encrypted storage with role-based permissions. All analyses reported in the paper use the de-identified extract.

Limitations of the dataset. The multi-country, school-based recruitment supports external validity across diverse contexts, yet the data remain observational and subject to participation and site-selection effects. As noted in the paper, labels are derived from clustering rather than external ground truth, and generalizability beyond the sampled populations should be evaluated in future work with prospective field validation. Below are the BIC clustering results.

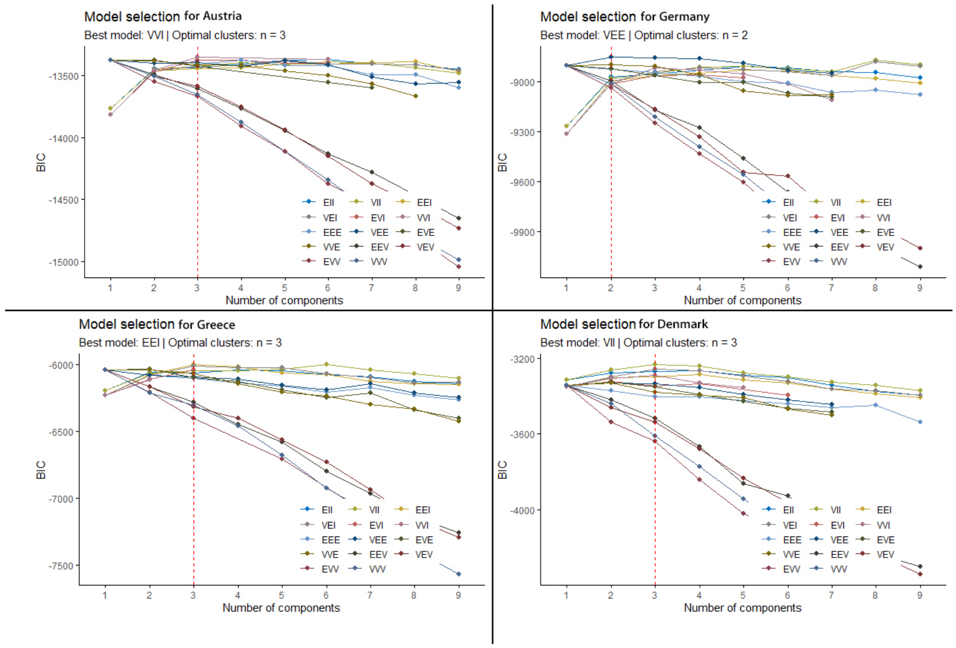

Figure A1: BIC Plots for Austria, Germany, Greece & Denmark Datasets

A.1.2 Hyperparameter Tuning and Cross-Validation Procedures

This section documents hyperparameter selection, cross-validation settings, and robustness procedures for all four country models (Austria, Germany, Greece, Denmark). The same pipeline was applied to each country unless explicitly noted. All analyses were conducted in Python using `scikit-learn`, `TensorFlow/Keras`, and `Stable-Baselines3`.

Random Forest (RF) Hyperparameter Tuning

RF hyperparameters were tuned separately per country using grid search on the training split (70%). The grid:

- `n_estimators` $\in \{200, 400, 600, 800, 1000\}$
- `max_depth` $\in \{6, 8, 10, 12, 14\}$
- `min_samples_split` $\in \{2, 3, 4, 5\}$
- `max_features` $\in \{\texttt{sqrt}, \texttt{log2}\}$

We used 5-fold stratified CV with weighted F1 as the scoring metric. The optimal region converged to:

$$\texttt{n\_estimators} = 1000, \quad \texttt{max\_depth} = 12, \quad \texttt{min\_samples\_split} = 3, \quad \texttt{max\_features} = \texttt{sqrt}.$$

These were fixed for downstream experiments. The selected RF served as (i) a standalone baseline and (ii) a component of the RF–DNN ensemble.

A.1.3 Deep Neural Network (DNN) Hyperparameter Tuning

DNNs were tuned with Keras Tuner `RandomSearch`. Search space:

- Hidden layer 1 units $\in \{32, 64, 96, 128, 160, 192, 224, 256\}$
- Hidden layer 2 units $\in \{16, 32, 48, 64, 80, 96, 112, 128\}$
- Dropout $\sim \mathcal{U}(0.1, 0.5)$
- Learning rate $\in \{10^{-2}, 10^{-3}, 10^{-4}\}$

Each trial trained up to 50 epochs with early stopping (patience 10, restore best weights), batch size 16, 10% validation split. We evaluated 10 configurations with two executions per trial and selected the highest validation accuracy. Convergence across countries:

$$\text{HL1} = 128\text{–}256,\ \text{HL2} = 32\text{–}64,\ \text{Dropout} \approx 0.2\text{–}0.3,\ \text{Optimizer=Adam},\ \eta = 10^{-3},\ \text{Output=3-way softmax.}$$

The chosen DNN was then retrained on the full training set per country for ensemble integration.

A.1.4 Ensemble Classifier (RF + DNN Soft Voting)

We combined RF and DNN class probabilities via weighted averaging. We tested:

$$\text{RF80–DNN20} \quad \text{and} \quad \text{RF60–DNN40.}$$

RF80–DNN20 consistently delivered the strongest hold-out performance in all countries and is reported as the main comparison model.

Feature Selection for Top-5 Models

Top-5 features were selected using country-specific RF Gini importances. Selected dimensions:

- **Austria**: Risk Awareness, Digital Literacy, Autonomy of Choice, Support Seeking, Regulation of Aggressive Impulses
- **Germany**: Respect Toward Others, Risk Awareness, Support Seeking, Digital Literacy, Autonomy within Digital Context
- **Greece**: Autonomy of Choice, Regulation of Aggressive Impulses, Risk Awareness, Support Seeking, Digital Literacy
- **Denmark**: Autonomy within Digital Context, Regulation of Aggressive Impulses, Digital Citizenship, Risk Awareness, Digital Literacy

The Top-5 subset was used to retrain RF, DNN, and their soft-voting ensemble.

Deep Reinforcement Learning (Hybrid DRL) Optimization

State representation: a compact 6D vector comprising the Top-3 RF-selected raw features and the DNN 3-class probability vector (knowledge distillation).

$$s = [x_{(i_1)}, x_{(i_2)}, x_{(i_3)}, P_{\mathrm{DNN}}(y{=}1), P_{\mathrm{DNN}}(y{=}2), P_{\mathrm{DNN}}(y{=}3)].$$

Algorithm and hyperparameters: Stable-Baselines3 DQN with

- Policy MLP: $[512, 256, 128]$ ReLU
- Learning rate $= 5 \times 10^{-5}$, buffer size = 100,000, batch size = 256
- Target network update every 500 steps
- Exploration fraction = 0.7, final $\epsilon = 0.01$
- Learning starts = 15,000 steps
- Total timesteps per country = 150,000–200,000

Values were held constant across countries for comparability.

Cross-Validation Procedures

1. 5-fold CV for RF, DNN, and Ensemble All baselines used 5-fold stratified CV on the training partition (70%), scored by weighted F1. Hyperparameters were tuned with CV and refit on the full training set.

2. 10-fold CV for DRL Given stochastic training and compute cost, we used 10-fold CV for the hybrid DRL. Each fold trained a fresh agent on the fold's training split. Validation Accuracy and F1 were averaged.

3. Paired t-tests We conducted paired t-tests over folds to compare:

RF80–DNN20 vs. Hybrid DRL and Top5 RF80–DNN20 vs. Hybrid DRL.

Example for Austria (F1): $t = -1.812$, $p = 0.144$. Hybrid DRL achieves comparable performance with fewer dimensions.

Robustness Checks

(1) Sensitivity analysis Learning rate $\in \{10^{-5}, 5{\times}10^{-5}, 10^{-4}\}$, batch size $\in \{64, 128, 256\}$. Performance was stable. Illustrative ranges:

- Austria: F1 $0.6077 \to 0.9287$
- Greece: F1 $0.5960 \to 0.8167$
- Denmark: F1 $0.4927 \to 0.8232$

Optima aligned with the chosen defaults (lr $= 5 \times 10^{-5}$, batch = 64–128).

(2) Multiple random seeds Five seeds $\in \{42, 123, 2024, 789, 456\}$. F1 standard deviation was low: ≈ 0.003–0.005 in Austria and Greece, 0.0000 in Denmark.

(3) Ablation (no DNN probabilities) Removing the DNN probability vector reduced F1 by: Austria -0.221, Germany -0.213, Greece -0.128, Denmark -0.178, demonstrating the value of distillation.

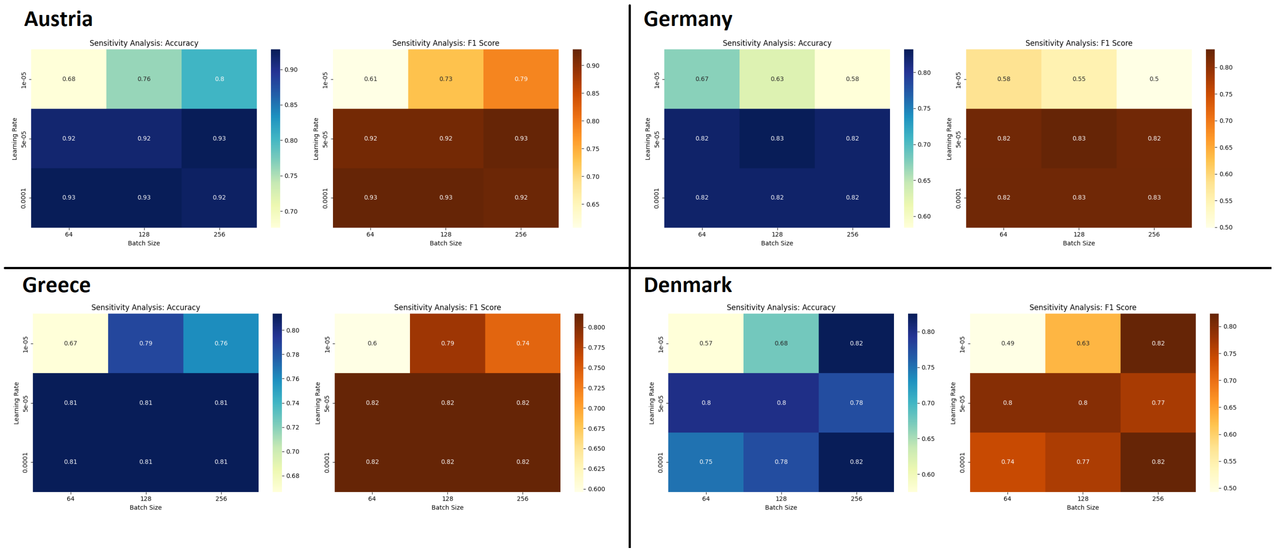

Figure A2: Robustness Check: Sensitivity Analysis

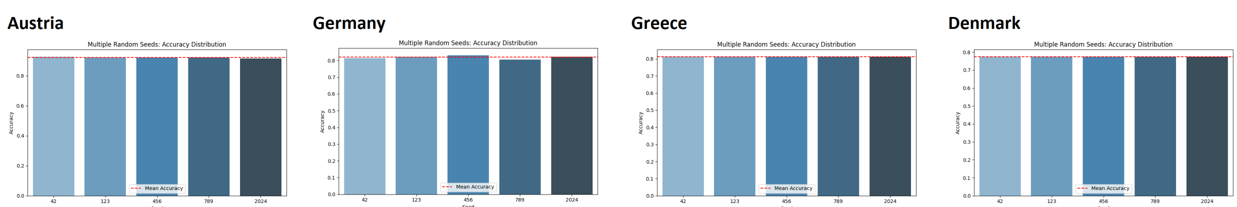

Figure A3: Robustness Check: Random Seed Accuracy

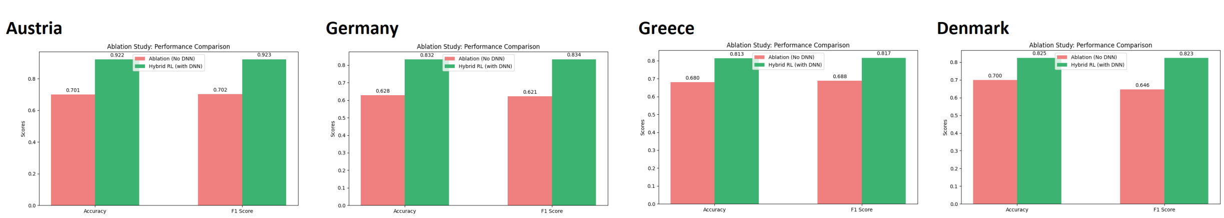

Figure A4: Robustness Check: Ablation Study

A.2 Appendix 2

A.2.1 Weekly Usage by Digital Maturity Terciles

Table A1 presents Week_Use summary statistics by maturity tercile for each subgroup. Figures B1 and B2 display boxplot distributions of Week_Use by maturity tercile within each Country × Gender × AgeBand combination.

Table A1: Weekly usage by maturity tercile and subgroup

Country	**Gender**	**Age group**	**Week_Use Low (M)**	**Week_Use High (M)**	**Δ (High–Low)**	**Cohen's d**
Austria	Female	11–14	3.81	3.30	−0.51	−0.69
Austria	Female	15–18	4.01	3.50	−0.52	−0.69
Austria	Male	11–14	3.71	3.36	−0.35	−0.47
Austria	Male	15–18	3.91	3.49	−0.42	−0.56
Germany	Female	11–14	3.56	3.19	−0.37	−0.49
Germany	Female	15–18	3.81	3.23	−0.58	−0.75
Germany	Male	11–14	3.54	3.39	−0.15	−0.19
Germany	Male	15–18	3.75	3.45	−0.30	−0.39

Note. M = mean; Δ = difference (High tercile minus Low tercile). Cohen's d computed as ($M_{\text{High}} - M_{\text{Low}}$) divided by pooled SD. Negative values indicate that higher-maturity adolescents report lower weekly usage.

Table A2: Weekly usage and age acceleration correlations

Country	**Gender**	**Age group**	**Spearman ρ**	**p-value**
Austria	Female	11–14	0.01	0.79
Austria	Female	15–18	0.01	0.92
Austria	Male	11–14	−0.05	0.22
Austria	Male	15–18	0.04	0.40
Germany	Female	11–14	−0.10	0.07
Germany	Female	15–18	−0.08	0.19
Germany	Male	11–14	−0.01	0.81
Germany	Male	15–18	0.06	0.30

Note. Spearman correlations between Week_Use and age acceleration (predicted age minus chronological age). Near-zero correlations indicate weekly usage does not predict the developmental signal captured by the maturity model.

While higher-maturity terciles tend toward slightly lower weekly usage in most subgroups (Table A1), this pattern is modest in magnitude and not uniform across contexts. More critically, weekly usage shows no meaningful correlation with the age acceleration signal (Table 2.5), indicating that the maturity model captures developmental variance independent of usage volume. These results suggest that digital maturity as operationalized through the DIMI reflects qualitative patterns of digital engagement rather than simply time spent online.

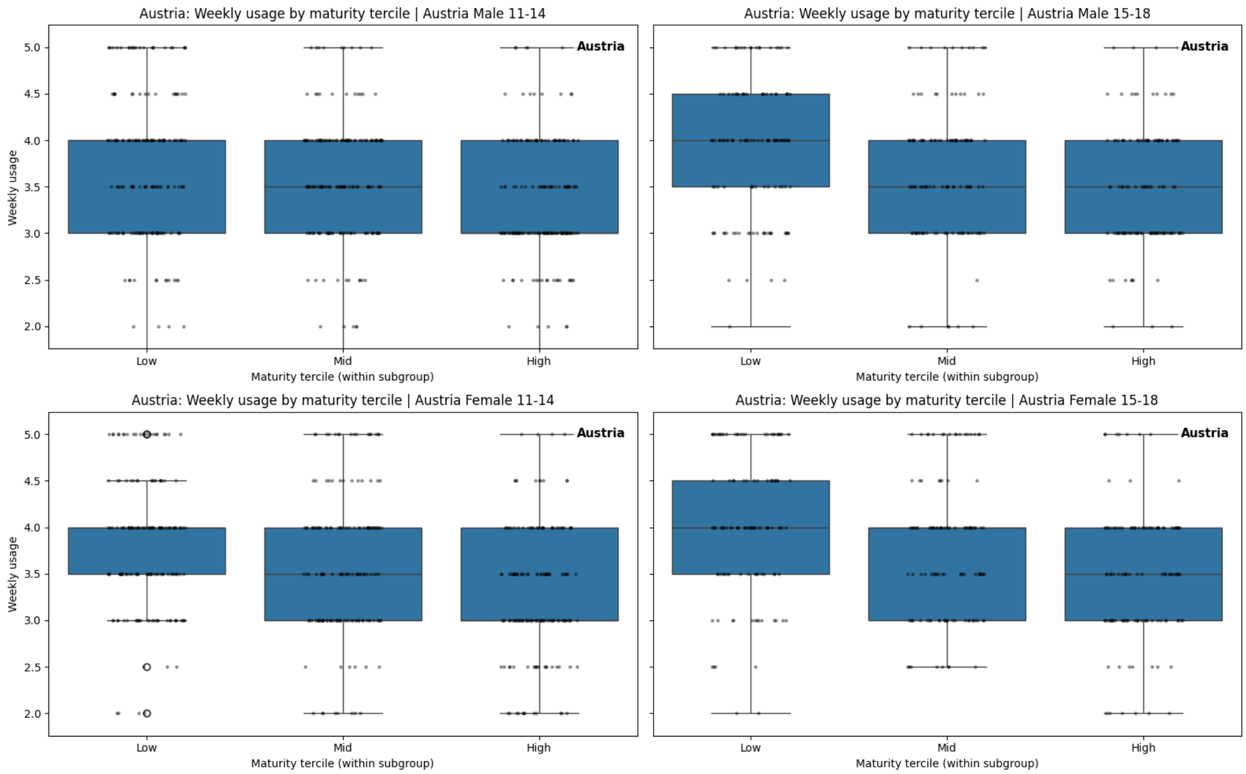

Figure A5: Austria Intensity of Usage Box Plot

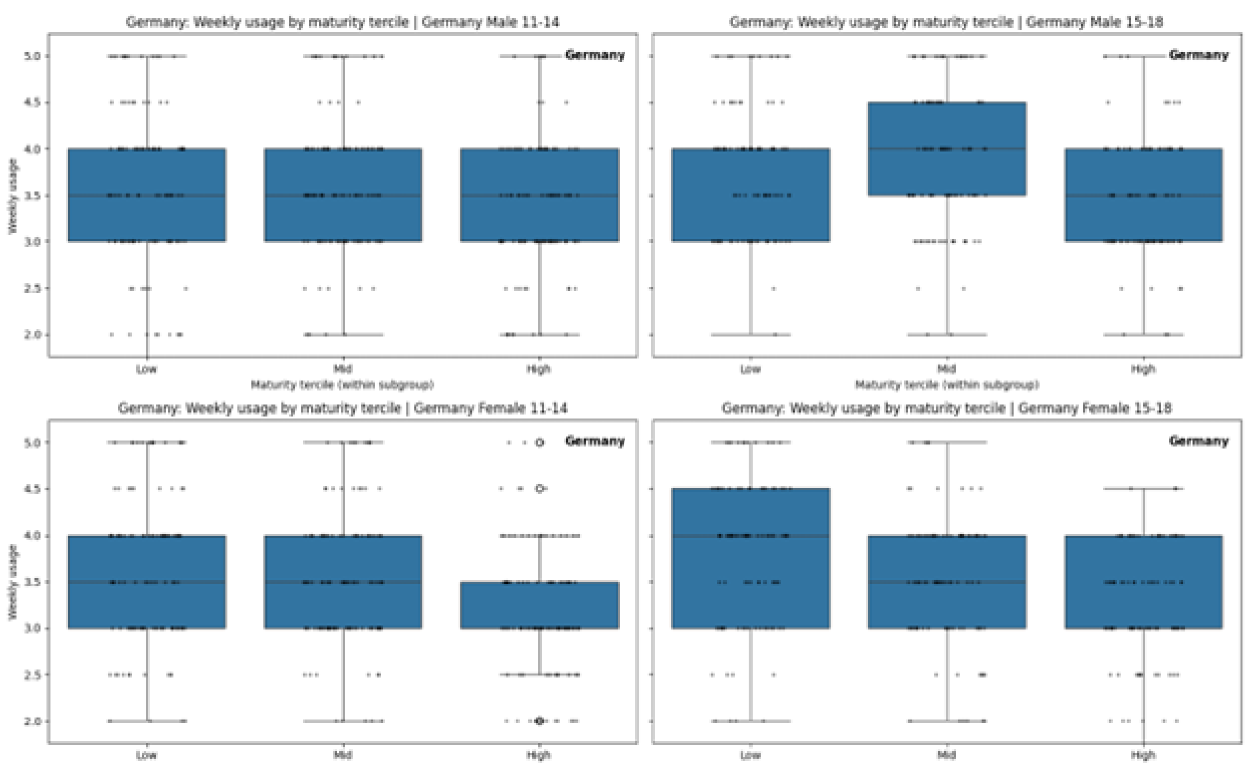

Figure A6: Germany Intensity of Usage Box Plot

A.3 Appendix 3

A.3.1 List of Stakeholders with Letter Reference Code

Stakeholder Group	Reference Code	Organisation Name
Academic/Research Institution	F2669466	Centre For Information Policy Leadership (CIPL)
Academic/Research Institution	F2669458	Institut Mines-Télécom Chaire Economie circulaire des données
Academic/Research Institution	F2669275	League of European Research Universities (LERU)
Academic/Research Institution	F2669199	Lund University
Academic/Research Institution	F2669381	Science Europe
Business Association	F2668404	American Chamber of Commerce to the EU (AmCham EU)
Business Association	F2669067	AMETIC
Business Association	F2669386	ANIMA CONFINDUSTRIA
Business Association	F2669469	Association for Financial Markets in Europe
Business Association	F2669465	Association of Commercial Television in Europe
Business Association	F2669442	Assuralia
Business Association	F2669443	BDEW German Association of Energy and Water Industries
Business Association	F2669371	BDI - Federation of German Industries
Business Association	F2669455	Bitkom e.V.
Business Association	F2669346	BSA - The Software Alliance
Business Association	F2669272	Bundesverband Digitale Wirtschaft (BVDW) e.V.
Business Association	F2669408	CCIA Europe (Computer & Communications Industry Association)
Business Association	F2669320	Cigref
Business Association	F2668337	Confartigianato Imprese
Business Association	F2669366	Confederation of Industry of the Czech Republic
Business Association	F2669437	Copa-Cogeca
Business Association	F2669244	Deutscher Industrie- und Handelskammertag e. V.
Business Association	F2669080	DIE FAMILIENUNTERNEHMER
Business Association	F2669226	DIGITALEUROPE
Business Association	F2669299	Eesti Infotehnoloogia ja Telekommunikatsiooni Liit
Business Association	F2669367	ETNO - European Telecommunications Network Operators' Association
Business Association	F2669270	EUROCHAMBRES - European Chambers of Commerce & Industry
Business Association	F2669400	European Association of Co-operative Banks
Business Association	F2669212	European Banking Federation
Business Association	F2669480	European Broadcasting Union
Business Association	F2669457	European Games Developer Federation
Business Association	F2669405	European Publishers Council
Business Association	F2669154	Federation of European Publishers (FEP-FEE)
Business Association	F2669242	FNSEA
Business Association	F2669235	French Publishers Association (SNE)
Business Association	F2636044	GANVAM, Spanish Association for Motor Trade, Repair and Parts
Business Association	F2669291	Independent Retail Europe

Stakeholder Group	File Name	Organisation Name
Business Association	F2669246	Interactive Software Federation of Europe
Business Association	F2669216	Lega Nazionale cooperative e mutue
Business Association	F2668364	Mouvement des Entreprises de France
Business Association	F2669358	News Media Europe
Business Association	F2669451	Numeum
Business Association	F2669327	Platform Economy/Plattformsföretagen
Business Association	F2669394	SIINDA
Business Association	F2669112	Technology Industries of Finland
Business Association	F2668972	The Federation of Finnish Enterprises
Business Association	F2668274	The Finnish Media Federation (Finnmedia)
Business Association	F2669482	U.S. Chamber of Commerce
Business Association	F2669379	Zentralverband des Deutschen Handwerks (ZDH)
Business Association	F2669426	Zentralverband Deutsches Kraftfahrzeuggewerbe e.V.
Company/Business organisation	F2669467	AFNUM (Association Française des Industries du Numérique)
Company/Business organisation	F2669309	Airbus
Company/Business organisation	F2669406	Belron International Limited
Company/Business organisation	F2669492	British Broadcasting Corporation
Company/Business organisation	F2669484	CONSEIL NATIONAL DES GREFFIERS DES TRIBUNAUX DE COMMERCE
Company/Business organisation	F2669357	Curve OS Limited
Company/Business organisation	F2669259	DATEV eG
Company/Business organisation	F2669280	DEKRA e.V.
Company/Business organisation	F2669298	Deutsche Börse Group
Company/Business organisation	F2669401	d-fine GmbH, a European consultancy and leader of the GAIA-X consortium EuroDaT (European Data Trustee) on behalf of Hessian Ministry of Economics, Energy, Transport, and Housing and Centre Responsible Digitality (ZEVEDI)
Company/Business organisation	F2669213	EnBW Energie Baden-Württemberg AG
Company/Business organisation	F2669489	Fujitsu
Company/Business organisation	F2669284	Data Sharing Coalition (DSC)
Company/Business organisation	F2665959	IBM
Company/Business organisation	F2669187	ING Bank N.V.
Company/Business organisation	F2669055	Ingka Group — IKEA
Company/Business organisation	F2669396	LKQ Europe Holdings / LKQ Europe GmbH
Company/Business organisation	F2669453	London Stock Exchange Group (LSEG)
Company/Business organisation	F2669111	MAIF
Company/Business organisation	F2669261	Mediaset S.p.A.
Company/Business organisation	F2669435	OVHcloud
Company/Business organisation	F2669459	Salesforce
Company/Business organisation	F2669358	Telefonica, S.A.
Company/Business organisation	F2660968	The Polish Confederation Lewiatan
Company/Business organisation	F2667291	Vodafone Group
Consumer Organisation	F2669324	Allgemeiner Deutscher Automobil-Club e.V. (ADAC e.V.)
Consumer Organisation	F2669325	Fédération Internationale de l'Automobile (FIA) Region I
EU Citizen	F2669271	-
NGO	F2663613	Allied for Startups

Stakeholder Group	File Name	Organisation Name
NGO	F2665964	Beltug, Belgian association of CIOs and Digital Technology leaders (490 organisations).
NGO	F2669288	COMMUNIA Association for the Public Domain
NGO	F2668966	Creative Commons
NGO	F2665975	Eurocities
NGO	F2669490	European Law Institute
NGO	F2669415	Open Future Foundation
Non-EU Citizen	F2667367	-
Other	F2669463	C4C
Other	F2669388	CSC - IT Center for Science
Other	F2669352	European Respiratory Society
Other	F2669468	Kapsch TrafficCom AG
Public Authority	F2669425	Danish Business Authority (DBA)
Public Authority	F2669364	European Data Protection Supervisor
Public Authority	F2669418	Finnish Transport and Communications Agency (Traficom)
Public Authority	F2669215	Hellenic Telecommunications and Post Commission (EETT)
Public Authority	F2669258	The Competition and Markets Authority (CMA)
Trade Union	F2669156	European Trade Union Confederation
Trade Union	F2669399	UNI Europa

A.3.2 VSD Value Classifier

Building the training dataset

For a high-quality training data, the authors constructed a comprehensive dataset comprising definitions and examples of the 12 VSD values from multiple academic sources like textbooks and research articles (Friedman et al., 2008, 2014, 2019). This formed the foundational knowledge for the classifier, allowing it to learn the fine distinctions and semantic dimensions inherent in each VSD value. Following procedures were conducted to further enhance the dataset:

- *Dataset enrichment*: To augment the depth and diversity of our dataset, we harnessed the capabilities of GPT-3.5, a large language model with an objective to enhance the dataset's richness. GPT-3.5 was employed to generate additional relevant content such as definitions, scenarios, and synonyms for each of the 12 VSD values. This process aimed to provide a more exhaustive and comprehensive representation of the subtle semantic dimensions within the VSD framework. The intention was to enrich the dataset with meaningful and contextually relevant variations, contributing to a more robust training process for the VSD Value Classifier.

- *Incorporation of RoBERTa*: To further strengthen the classifier's adaptability and robustness, and the need for a more exhaustive understanding of the intricate language variations and contextual granularities within legal texts, led to the use of RoBERTa (Robustly optimized BERT approach). BERT (Bidirectional Encoder Representations from Transformers) is a large language model widely used for natural language processing tasks. It excels at capturing contextual information and understanding complex language variations. Building upon BERT, RoBERTa further refines the pre-training process by introducing dynamic patterns, enabling it to capture more dependencies and contextual information over longer pieces of text. This modification makes RoBERTa particularly effective in domains with challenges such as the legal domain, where distinct vocabulary and contextual variations demand a more diligent approach. In this research, RoBERTa plays a pivotal role in enhancing the training dataset. It generates variations of the definitions of VSD values already present in the dataset. These variations introduce subtleties, ensuring that the model can handle a diverse range of textual details and complexities inherent in legal language. By enriching the dataset with RoBERTa-generated variations, the authors aim to fortify the classifier's ability to discern and classify complex language patterns in legal texts more effectively.

The integration of RoBERTa complements the utilization of GPT-3.5, the initial language model used in our study. While GPT-3.5 excels in generating human-like text, RoBERTa's focus on contextual understanding and variations makes it a valuable addition to our methodology. Combining the outputs of these two models create a more robust and adaptable classifier, ensuring a comprehensive coverage of the complicated language landscape within legal texts. This dual-model approach enhances the overall efficacy of the training process, contributing to the classification accuracy of the VSD values. The final training dataset consisted of 1320 entries with a total of 49848 words, encompassing 110 variations for each VSD dimension.

Use of Pre-Trained Model

In the domain of NLP, a transformative approach centering around the use of Pre-trained Models (PTMs) has gained considerable traction across academic and industry landscapes. These models are pre-trained on a large scale corpora of text for either word prediction, next sentence prediction or is made to learn dependencies and contextual information over longer pieces of text. Then this pre-trained model is fine-tuned on the texts of a specific domain, targeting a downstream task of that domain (Song et al., 2022a). This pre-training phase equips the model with a general understanding of language structures. Within the legal domain, where challenges such as the scarcity of large-scale, high-quality datasets and distinctive vocabulary prevail, the effectiveness of Pre-trained Language Models (PLMs) has been notably showcased across various tasks (Song et al., 2022a). These tasks include binary classifications (Lippi et al., 2019; Zheng et al., 2021), multi-label classifications (Song et al., 2022b; Chalkidis et al., 2019), summarizing (Gargett et al., 2020; Kornilova and Eidelman, 2019), and information retrieval (Rabelo et al., 2022), demonstrating the versatile applicability of PLMs in legal contexts.

In line with this paradigm, our study harnesses the power of a state-of-the-art pre-trained model known as EUBERT for the specific task of classifying paragraphs into the 12 values proposed by the VSD framework. EUBERT, a fine-tuned BERT uncased model, stands out as an exemplar in this domain. Initially pre-trained on a substantial corpus of documents compiled by the European Publications Office over the past three decades (Campion, 2023), EUBERT offers a unique advantage in understanding the intricacies of policy documents and stakeholder recommendations. To further enhance EUBERT's capabilities and tailor it specifically for our research objectives, we fine-tune the model on our dataset. This advanced training process enables EUBERT to capture domain-specific intricacies present in legal and policy-oriented language. The paramount significance of training EUBERT on our dataset lies in its augmented ability to discern subtle semantic meanings, ensuring a more advanced and accurate text classification within the VSD framework. This refined methodological approach serves as a powerful tool, providing depth and precision in decoding convulated textual data. As we delve into the systematic exploration and alignment of stakeholder values with policy objectives outlined in the EU Data Act, the advanced version of EUBERT becomes instrumental in unraveling the complexities embedded in legal and policy-oriented language.

Training the Model

To facilitate the training of EUBERT to get our VSD Value Classifier, we employed the Hugging Face platform, a renowned and user-friendly platform for NLP practitioners. Hugging Face provides a collaborative space where researchers and developers can seamlessly access, share, and fine-tune various pre-trained models for their specific tasks. This classifier, trained with precision on the Hugging Face platform, serves as the fundamental element in our methodology, enabling the classification of paragraphs into the 12 values proposed by the VSD framework. The training and testing data was divided into an 80%-20% split, approximately 39,878 words for training the EUBERT model and 9,965 words for testing its accuracy. The classifier attains a notable accuracy of 98.1%.

Validation of the Classifier

To address the concerns regarding the validation of the VSD classifier and enhance the general validity of our analysis and findings, we undertook an additional validation procedure. This involved testing the classifier on a text corpus consisting of randomly generated paragraphs as well as selected paragraphs from the European Journal of Information Systems (EJIS) website. We also show the classification of the first Article in the Draft EU Data Act that contains 3 paragraphs. The goal was to demonstrate the classifier's performance on text fragments that do not necessarily contain any value statements and to provide more transparency and summary statistics about the classification outcomes.

Random Paragraphs Validation: We generated several random paragraphs using an online tool (https://randomwordgenerator.com/paragraph.php) and classified them using our VSD classifier. The classifier assigned the most probable VSD value to each paragraph, along with a probability score. As expected, the probability scores for these random paragraphs were consistently low (below 50%), indicating that the classifier could recognize the lack of clear value statements in these texts.

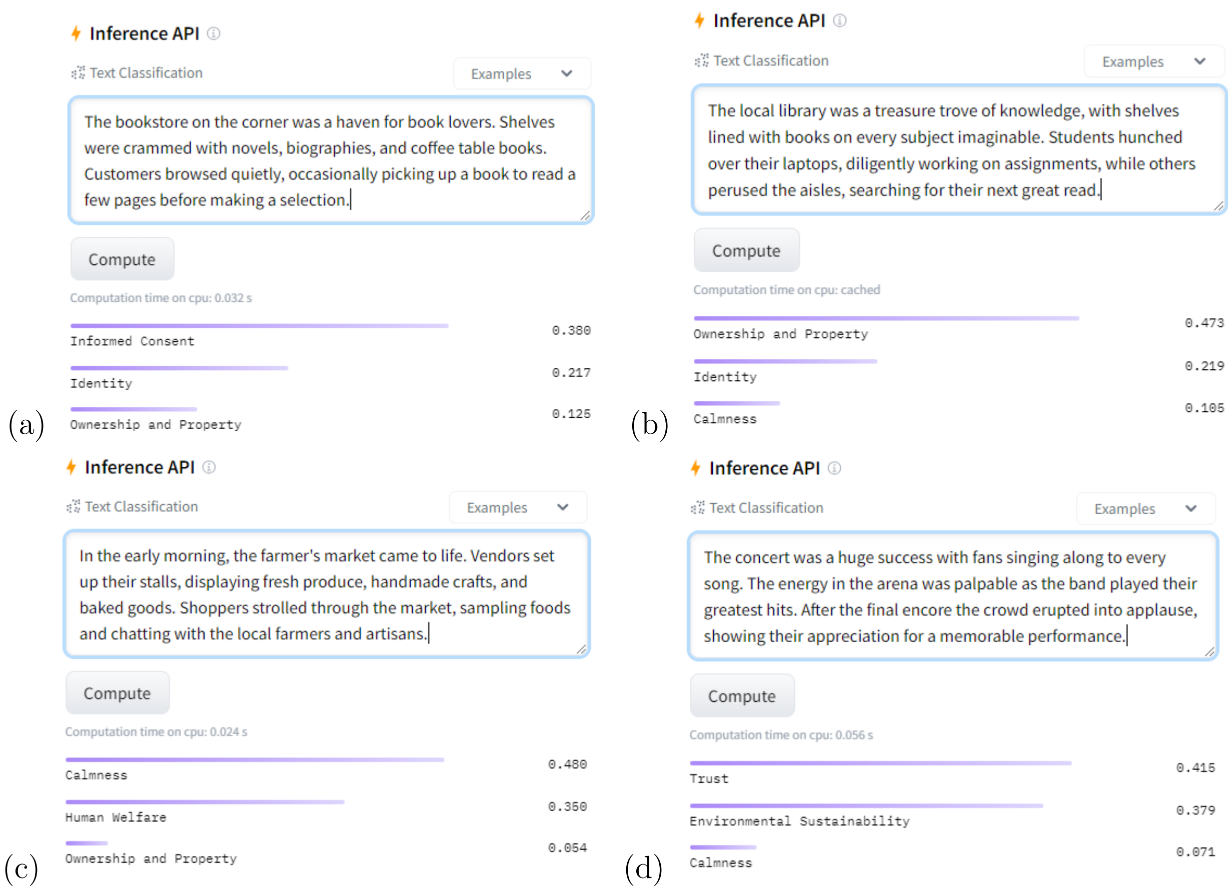

Figure A7: Classification Probabilities for Randomly Generated Paragraphs