Summer 2021
AI Policy Research Clinic

Policy Paper

Working Group 1

Bruna de Castro e Silva
Erika Ly
Franziska Poszler
G. R. Marvez
Nagadivya Balasubramaniam
Thomas Vogl

Municipal Stakeholder Engagement Strategies for Learning Analytics and AI in Education:

Participatory Design, Accountability and Oversight Mechanisms

July 2021

Bruna de Castro e Silva, Erika Ly, Franziska Poszler,
G. R. Marvez, Nagadivya Balasubramaniam, Thomas Vogl

This report was produced in the Summer of 2021 as part of a Research Clinic led by the Berkman Klein Center's Ethics of Digitalization initiative, in collaboration with the City of Helsinki, their partner Saidot, the Global Network of Internet & Society Centers, and under the patronage of the Federal President of Germany Frank-Walter Steinmeier.

The aim of the Clinic was to translate best practices and principles concerning the use of AI by public service providers into actionable measures, applied within the context of municipal educational programs. We worked with participants and experts from many disciplines and from across the globe, underpinned by an ethos of interdisciplinary collaboration, mutual learning, and open exchange.

Contents

List of Tables and Figures

1 Introduction

1.1 Research Background

Artificial Intelligence and machine-learning bring transformational promise for education and learning. Against the backdrop of the COVID-19 pandemic, students and teachers around the world have relied on technology to open virtual classrooms, uptake remote learning, and deliver continued education. While we have seen meaningful benefits, there is a greater opportunity to benefit from an AI-powered education system.

The Education Division of the City of Helsinki is exploring AI and machine-learning solutions to improve pedagogical and social outcomes for students and teachers. In addition to reducing administrative and operational burdens for teachers and schools, AI supported-learning analytics tools can improve experiences through personalizing student learning, promote the completion of studies through automated notifications, leverage predictive models to identify timely interventions for at-risk learners, and provide greater insights for teachers into student competencies, well-being, and learning outcomes.

These opportunities are not without risk. The use of data, coupled with explanatory and predictive modelling of human behavior is an inherently complex enterprise and any use of predictive analytics to inform decisions affecting individuals carries significant ethical, legal, and social obligations.

1.2 Research Clinic

The Summer 2021 AI Policy Clinic Sprint, hosted by the Berkman Klein Center for Internet and Society at Harvard University, in collaboration with the Education Division of the City of Helsinki, their partner, Saidot, and the Global Network of Internet & Society Centers, was intended to translate best practices and principles concerning AI technologies into actionable measures for public-sector educational and vocational programs.

The goal of the clinic was to assist local governments to develop an ethical governance and stakeholder engagement model for the responsible use of AI solutions in educational settings, with a focus on participatory design as well as accountability and oversight.

The purpose was to assist policymakers and other high-level stakeholders to understand key risks, opportunities, and implications of analytics and AI in education; while providing practical strategies for participatory design, and mechanisms for accountability and oversight. Specifically, the clinic was intended to

1. Assist the City of Helsinki’s Education Division to create an inclusive, participatory, and sustainable strategy for stakeholder engagement throughout the design, development, deployment, and assessment stages of new classroom technologies; and

2. Assist the City of Helsinki's Education Division to develop accountability and transparency governance frameworks and practical implementation strategies.

Finally, the specific objectives in Helsinki for this project were related to providing all citizens with equal opportunities and promoting learners' development and wellbeing as well as lifelong learning. These objectives come from the value base of Finland's education system, which can be summarized as: growth into humanity, the unique value of the individual learner, equality and diversity, sustainability, and rights of the child.

1.3 Research Summary

1.3.1 Research Process

Over a period of three weeks, the interdisciplinary project team gathered relevant insights about the case (including the specific use case in vocational education provided by the City of Helsinki), conducted a scan of the relevant literature, consulted existing frameworks, engaged with experts, and synthesized knowledge to produce outputs relevant to the clinic's public sector partner (the City of Helsinki's Education Division), and, more generally, for public administrators pursuing their own AI projects and seeking transferable knowledge.

1.3.2 Summary of Findings

The key findings from the research in response to the City of Helsinki's use case involved a pragmatic adaptation of high-level principles and guidelines into actionable approaches that the City could take, given its context.

1. In terms of participation, engagement with stakeholders should be seen as more than an event, but instead as a long-term iterative process, and one that occurs across multiple stakeholder contexts. To this end, the team produced a risk matrix linking high-level human rights principles to risks in this setting, relevant stakeholders, and a menu of participatory approaches. While not exhaustive, it provides a framework within which to think systematically about how to create meaningful opportunities for participation across different elements of the project and at different stages of its development. The team also produced a set of website mock-ups to encourage the City to build participation into its AI register to enable interested stakeholders to gather information about how to get involved and leave comments on the initiative and the participatory process.

2. In terms of accountability and oversight, the focus was on developing a specific mechanism for human control and oversight and describing what steps should be followed when trying to produce a meaningful accountability mechanism such as this. To this end, the team has proposed a model that includes a steering committee and working groups, made up of relevant internal, external, and cross-functional membership. These bodies are responsible for translating ethical and human rights principles into both high-level decisions and the

practical applications, such as learning analytics and AI tools. A Human Oversight Translation document was developed to help support this translation work between governance, socio-technical, and technical layers. The team also produced a set of website mock-ups to encourage the City to build human oversight into its AI register to enable oversight bodies to monitor and evaluate the progress of learning analytics and AI tools.

1.3.3 Concluding Remarks

Our approach includes four key linking heuristics that help align participation and human oversight:

1. The technology project lifecycle (from exploratory stages, through design and development, to ongoing maintenance and evaluation).
2. The layers of organizational activity (from technical, through socio-technical/operational, to governance).
3. An ethical and human rights framework (based on EU guidance and law); and
4. A website tool (that builds participation and accountability into existing infrastructure).

Our model takes what is currently in place in Helsinki and adds meaningful practical and theoretical layers that tie together participation and human oversight:

- Helsinki has design staff on their team – we are giving them tools to think about participation across the lifecycle, in terms of different layers of project governance, and through an ethics and human rights approach.
- Helsinki has an existing governance structure in place but faces challenges when trying to get members of different functional groups to understand each other – this is why we are recommending interdisciplinary teams with diverse (internal, external, and cross-functional) membership to build mutual understanding, facilitate translation, and enable participation in governance and human oversight.
- Helsinki has an existing AI register – we have recommended new features and functionality to support participation and human oversight.

2 Definitions and Methodology

2.1 Key Governance Terms

2.1.1 Participatory Design

Participatory design is an approach to co-creating products or services with relevant stakeholders at all stages of a design process. Participatory methods aim to create alignment and understanding between collaborators from different backgrounds to assist in problem identification, information gathering, as well as the design of the output. They may also be used to support the integration of interdisciplinary participation with stakeholder engagement (O'Brien et al., 2013), which could assist with building understanding on teams with members that have different backgrounds, functions, and worldviews.

Importantly, participation is not a singular event but a continuous process that occurs in different modes (e.g., gathering feedback, deliberation, user experience testing) (Brereton & Buur, 2008; Sanders et al., 2010) as stand-alone participatory methods have limitations and may only be useful in certain circumstances. Each mode of participation embodies a unique actor and sensor. By using different modes and applying them across the design process, it is possible to identify and resolve unanticipated problems before they become catastrophic.

2.1.2 Accountability

Accountability is a key concept spanning several domains including political science, business management, and social psychology. For this public sector education context, a definition of accountability also considers qualities of fairness and equity of the governance structures in which citizens are primary stakeholders. As such, the definition adopted here is: '*[a]ccountability is a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgement, and the actor may face consequences.'* (Bovens, 2007, p. 450).

It should also be noted that from a broader perspective, accountability is the root for concepts such as transparency, equity, democracy, efficiency, responsiveness, responsibility, and integrity (Verdiesen et al., 2021). Furthermore, auditing and algorithmic culpability are highlighted in the literature for the purpose of AI accountability (Gualdi and Cordella, 2021).

2.1.3 Oversight

Human oversight plays a role in both participation and accountability. To support this dual role we have recommended the creation of a steering committee and working groups. In terms of participation and engagement, the steering committee and working groups play a role in translating project objectives to the steering committee to senior leadership and the working groups to ground-level design and development teams. In terms of accountability, the steering committee and working groups will each look at different outputs as the project progresses to provide ethical governance and enable stakeholder engagement (for more details on what types of outputs these groups may engage with, see the detailed High-Level Model for Human Oversight).

Focusing on learning outcomes data alone may not be sufficient as an approach to governance (Lassnigg, 2012), but through AI project focused steering committees and project groups, the knowledge articulation needed to enable ethical governance and stakeholder engagement can be achieved (Brunet, 2019; Brunet & Aubry, 2018).

2.2 Key Technical Terms

2.2.1 Learning Analytics

The history of AI in education traces back to the 1970s with discussions about computer assisted instruction (Dreyfus et al., 1986; Dwyer, 1974). From these early days, there were debates about how best to use these technologies, not only to drill and practice rules and facts, but also to teach students how to learn (Dwyer, 1974). Similarly, there were debates about how intelligent computer assisted instruction would reshape the relationship between students, teachers, and technology, as well as what ethical implications these changes would have (Croy, 1989). Since then, the field of learning analytics has evolved into providing the technical foundation for intelligent learning environments. Learning analytics deals with the collection, measurement, analysis, and reporting of learner data and their pedagogical context with the intent to improve or optimize learning within the environment in which it occurs (Lang, Siemens, Wise & Gašević, 2017). Types of data gathered may range from insights provided from the use of learning tools and platforms, behavioral patterns across leaner journeys, or sensor data from the surrounding learning environment. Moreover, learning analytics can range from descriptive, diagnostic, predictive, or prescriptive, with each category leading to different decision-making sets. For example, in MOOCs, diagnostic analytics may assist in creating adaptive course content modified to best support a learner's needs, while predictive student data in MOOCs has been used to build classifiers to predict dropout and build a survey that can bring students back into the course materials (Whitehill et al., 2015). On a more negative side, intelligent tutoring systems are often "gamed" by students who try to complete course content as fast as possible (Baker, 2004; Walonoski & Heffernan, 2006).

2.2.2 Artificial Intelligence

Since research on AI began in the 1950s and 1960s, there have been two main approaches: the first wave focused on rule following algorithms and the second on self-learning pattern recognition (Dreyfus & Dreyfus, 1991; Dreyfus, 2007). Today, artificial intelligence is typically understood to be the ability of a computer to perform cognitive tasks associated with the capabilities of the human mind—this includes sensing, acting, problem-solving, and decision-making. At a broader level, AI is a field of study and practice that combines computer science and engineering, with datasets and algorithms, to solve problems through predictive algorithms that were thought to only be solved by human intervention. Some example technologies include computer vision and recognition, machine learning, as well as robotics and autonomous systems. While some continue to believe that AI replicates human intelligence (Krafft et al., 2020), the definition adopted for the purposes of this work aligns itself to 'narrow' or 'weak' AI in which programs are trained and adapted to perform specific tasks within the context of public education and learning.

2.2.3 Cyber-Physical System

The term 'cyber-physical system' (CPS) is attributed to Dr Helen Gill of the American National Science Foundation and understood to be physical and engineered systems whose operations are 'integrated, monitored, and controlled by a computational core' (Gill, 2008). More recently, CPSs have featured as a cornerstone concept for national innovation strategies focused on the Fourth Industrial Revolution or "Industry 4.0". Adopted for this report, CPSs are systems within user defined boundaries that tightly integrate cyber components (networked, computational, communication elements) and physical components (hardware, sensors, human operators) within a particular socio-technical environment. CPSs are characterized by feedback loops, co-evolution with the components contained within them, as well as integration or interdependence with other systems (Lee 2007; Ackoff, 1971). For example, applied to this context, a cyber-physical system for education would consider a networked classroom enabled by IoT sensors with ambient computing functions, underpinned by a technical infrastructure and human-interfaces, and includes teachers and students as co-actors or components within the physical system that provide data for the CPS to analyze, affect, adapt and evolve.

One potential challenge in CPSs is related to the dominant ways of thinking in different but interacting stakeholder groups. Two groups, such as managers and professionals, may have competing logics that inform their interests and activities (Reay & Hinings, 2009). The risk is that when one perspective dominates, it inspires resistance in the group holding the other perspective, for example when performance management is privileged by managers and IT developers over the expertise and quality of service privileged by professionals (White et al., 2010). The challenge is to find a balance between logics, such that the informational needs of one group, for example managers, do not interfere with the informational needs of another and allow for a melding of technology and human expertise that can augment teaching (Dreyfus et al., 1986; Dwyer, 1974), inform the organizational environment and allow workers to engage their intellective skills (Zuboff, 1985, 1988), or support artificing between public sector workers and algorithmic decision tools (Snow, 2021). For Helsinki to realize the benefits of AI and learning analytics, it will need to pay attention to these factors of competing institutional logics and come to some type of consensus that makes the CPS work for everyone involved. It is by means of a meaningful participatory approach with all affected stakeholders and an intelligible system of accountability, that these objectives can be achieved.

2.3 Technology in Education

AI in education is motivated to improve students' learning possibilities and their lifelong opportunities (Fadel et al., 2019). Studies relating to the testing of AI in the education sector have supported that AI can help improve learning opportunities for students and provide information to management systems. Also, AI technologies aim to ensure inclusive, equitable, and quality education (UNESCO Institute for Lifelong Learning, 2019). Moreover, it caters to learning opportunities for marginalized communities, people with disabilities, refugees, and school dropouts.

One of the transformative aspects of AI in education is its ability to create personalized learning. It can also support teachers to focus more on the students with learning difficulties as AI eases the teacher's regular mundane administrative tasks, e.g., AI as an assessment tool (UNESCO Institute for Lifelong Learning, 2019). Personalized education content can be designed based on the individual needs of students, and it supports their study performance (Shalini & Tewari, 2020).

AI and learning analytics in education pose some risks and ethical issues. One of the key issues is the privacy of the data which includes consent on the data use, data interpretation, and data management practices (Fadel et al., 2019). One possible example of the Data Pipeline Model for AI-HOKS with a special focus on student data privacy is illustrated below:

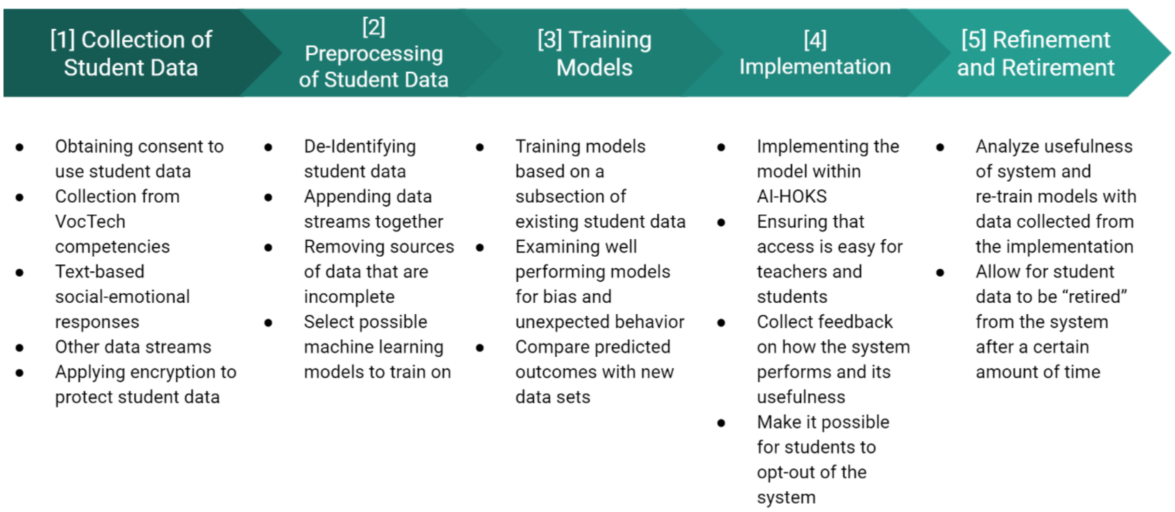


Figure 1: Data Pipeline Model for AI-HOKS with a Special Focus on Student Data Privacy.

The risk of data misuse is portrayed as a prime risk relating to AI in education. As an emerging field, the scope of AI in education is still narrow. In addition, the user experience is highlighted where agency, transparency, and intelligibility play a role when adopting AI for education (Fadel et al., 2019). Finally, an ethics of AI in education approach highlights potential ethical issues relating to consent, transparency, control over data, and algorithmic bias.

2.4 Methodology

This report's methodology focuses on interdisciplinary methods that include human-rights based, applied, visual, grounded, and translational approaches. By examining relevant literature, policy documents, and EU law, we were able to determine ways in which participatory design and human oversight could be integrated into the AI-HOKS implementation. It is at the intersection of policy and ethics, participatory design, and AI in education where we have found the methods and frameworks of best practices for AI design.

Through holistic literature review and consultations with the Helsinki team, we landed on the following qualities of our outputs: practical implementations, citizen access to data, AI ethics, human rights, and extendable to broader AI applications in different sectors. We also discovered gaps in communication between teams of differing expertise, such as teachers, developers, and government officials, and strived to develop translational models on high-level AI ethics and human oversight requirements from a technical and legal perspective.

2.4.1 Bottom-up and top-down approaches

The method that our group adopted to generate advice about how Helsinki should move forward with participation and human oversight in their learning analytics and AI projects was to meld bottom-up and top-down approaches.

By applying a bottom-up approach, we gained insights into the contextual factors that shaped the case and the foundational, organizational, and informational infrastructure that the City of Helsinki already has in place. This included understanding the existing stakeholders, the organizational structure, context surrounding the project, the available resources, partnerships, and information technology infrastructure. From there we discussed what kinds of practical frameworks could be introduced to this environment, including around how to understand the project lifecycle, the types of participation that would be most relevant, and the types of governance bodies that could best support human oversight. We explored and synthesized the literature to identify those evidence-based approaches that appeared to have the best fit with Helsinki's context.

The contextual features and practical approaches adopted are consistent with some elements found in examples in other jurisdictions (e.g., Darwin, Singapore), which lends further support to the approaches that we adapted to meet Helsinki's unique context. . While the recommendations may be unique to Helsinki, the framework underlying those recommendations could serve as a playbook for other jurisdictions hoping to learn from this case and enhance their own AI adoption efforts.

We also adopted a top-down ethics and human rights-based approach. The human rights-based approach to understand and address the societal risks of AI draws holistically on international human rights (Fukuda-Parr & Gibbons, 2021). It is a more robust framework than ethics of AI because human rights are internationally agreed norms, standards, and principles, established on international human rights law, and enforceable in international, regional, and national systems.

Although ethical guidelines for AI are voluntary, they also offer a valuable framework of ethical requirements and principles which contribute to the development of public policy and regulations concerning AI and other emerging technologies. Also, the recent advancement and promotion of AI ethics frameworks endorsed by powerful public and private actors in the field of AI have permitted ethics to play a significant role in AI governance.

In this work, high-level ethical principles for AI and human rights-based approaches were used as lenses to address the complex and multifaceted challenge of ethical AI governance in the public sector. These two top-down approaches were used not only as goal-oriented frameworks that informed the problem framing, research strategy, and methods; but also, as part of the theoretical foundation. Since one of the main elements of this research is multidisciplinarity, AI ethics and human rights were combined with other disciplines (e.g., computer sciences; design; educational sciences) in order to better understand and propose actionable and practical solutions for an ethical governance of AI and learning analytics at Vocational Education and Training (VET) in Helsinki.

2.4.2 Constraints and Limitations

The constraint of the bottom-up approach refers to the particularities of the concrete use case *per se* and how it differentiates from other international cases. As we were trying to align the approach with Helsinki's context, the lessons learned and best practices that we have transferred required adaptation and translation to the concrete case. This means that some recommendations may or may not be appropriate to Helsinki's practical context and we encourage the City to monitor the elements of this approach for effectiveness and unanticipated consequences.

The limitations of the top-down approach based on high-level ethical principles and human rights can be described as follows. First, sometimes two or more ethical principles or rights can collide with each other in a specific concrete case. These types of conflicts are extremely hard to solve, where even more abstract criteria are often recommended (e.g., the principle of proportionality) which can make the operationalization of ethical principles even more challenging in practical terms.

Second, the high level of abstraction of AI ethical principles can make their effective implementation difficult, and, consequently, there is a risk of this type of top-down approach becoming limited to vague terminology that lacks concrete enforcement mechanisms and “realistic means for shaping the design, development, and deployment of AI in the real world.” (Fukuda-Parr & Gibbons, 2021, p. 11).

3 Key Stakeholders in Education

3.1 Primary Stakeholders

3.1.1 Students

Students are one of the cornerstones of educational technology development. They are often the primary actor within a new system or curriculum. Thus, their voices in the development of new processes are critical to a tool that engages students and meets their needs. Without their input, developers risk missing what tools students actually use and could be unaware of the needs of students. Additionally, when students act as stakeholders, developers must provide an extra layer of security when processing their data. For VET students, they should give inputs on what kinds of assistance they would like to receive, should the tool determine that they are at risk for dropping out.

In design workshops with students, designers may find it helpful to ask the following questions:

1. How would you define learning success for yourself?
2. What kind of supports would be helpful to you personally to ensure timely graduation?
3. What is the best way for you to check in with your teachers about your academic progress?

3.1.2 Schools and Teachers

In stakeholder conversations, teachers should be looked towards for answers on what kinds of data are useful to them in making assessment decisions and how they provide social-emotional support to students. In the development of new education technologies, it is critical that educational theory and pedagogy is applied, and unaware computer scientists may rely on schools to help provide this framework. Without the input of teachers, designers risk misunderstanding the school-based context and how a new technology could be integrated into it.

For working with schools and teachers on the design of the AI system, it may be helpful to ask them the following kinds of questions:

1. What kinds of data about students and learning is the most useful to you?
2. Is there anything you wish you could change about the way student data is organized?
3. How do you currently support student success and how do you think such an AI system could fit into your current practice?
4. What characteristics of students have you noticed make them successful in their apprenticeships?

3.1.3 Technologists

Technologists include the computer scientists, designers, and machine learning engineers of the educational AI system. In working with these stakeholders, it will be important to map out how student data will be secured, the pipeline of data processing, and how teachers and students can access the data. Critically, technologists need to be able to explain their designs to non-technical stakeholders in understandable terms, so that they can receive feedback from teachers, students, and policymakers on how data streams and access portals are constructed. Additionally, these stakeholders can provide insight on the best way to make data processing transparent for the everyday user of the AI system.

Though technologists are the primary designers of the system, it may still be useful to ask the following:

1. What platforms for data storage or user interaction do you see this technology working best on?
2. What student data would be most useful in building a predictive model?
3. How do you plan to secure student data?

3.1.4 Policymakers

Policymakers provide insight into the legal and security issues that surround working with student data in an AI system. By working with technologists, they can help ensure that student data is treated correctly by identifying to users how their data is used and giving possible options for opting out of AI data analytics. Additionally, policymakers will be instrumental in updating Finland's procedures for AI systems should the EU's AI Act pass, as it will classify the AI-HOKS system as high-risk.

Policymakers are concerned with protecting student data rights, and it may be helpful to ask the following:

1. What concerns do you have about AI that interacts with student data?
2. How can you help us ensure that the system we build conforms to national and EU law?
3. What ethical or human oversight concerns might we be ignoring in our designs?

3.2 Secondary Stakeholders

Identifying other stakeholders within the community helps inform policy and technology development, as well as priorities for implementation. Differing perspectives from key stakeholders also assists with gaining buy-in to the eventual integration of the system By mapping out all the actors that may be users or indirectly related to the system, a stronger stakeholder or engagement community can be established for the system design.

- <u>Educational Leaders:</u> Directors of Education, Vocational Education, Executive Director
- <u>National Offices:</u> Education, auditing, and curriculum setting offices
- <u>Advocacy and Interest Bodies:</u> Human Rights, Child and Youth Groups
- <u>Group Associations:</u> Teachers unions and associations, employer associations, boards of trade, parent associations, student unions
- <u>Other public-service sectors</u>: mental health and physical wellness, youth programming, child and family services, housing, benefits and welfare
- <u>Other local governments</u>*:* for mutual learning and sharing best practice

4 Model for Stakeholder Engagement

4.1 Critical Thematic Questions

Building on the sections scaffolded above and in combination and in line with the clinic themes, two sets of questions were developed in order to create a model for stakeholder engagement and human oversight.

Critical questions for participation	*Where does participation happen?*
	What is the purpose of the participation?
	What are the meanings and insights from the participation?
	Do you want to host some participatory events, or embed participation in every aspect of the project?
Critical questions for Human Oversight and Accountability	*How is human oversight structured?*
	Who is responsible for what at what stage?
	What are the lines of responsibility and what are the materials and evidence that underpin accountability assessments?
	What does human oversight and accountability achieve?

Table 1: Critical Questions for Model of Stakeholder Engagement and Human Oversight.

4.2 Model Development for Stakeholder Engagement

The proposed model for stakeholder engagement draws from cross-disciplinary literature and practice. It takes into account the Technology Development Cycle and combines this with Organizational Design and Management frameworks, with interventions from Participatory Design, Accountability and Oversight which are in turn, informed by research and best practice frameworks. The model also aligns with Helsinki's input provided throughout the research clinic and the existing approach to participatory co-development.

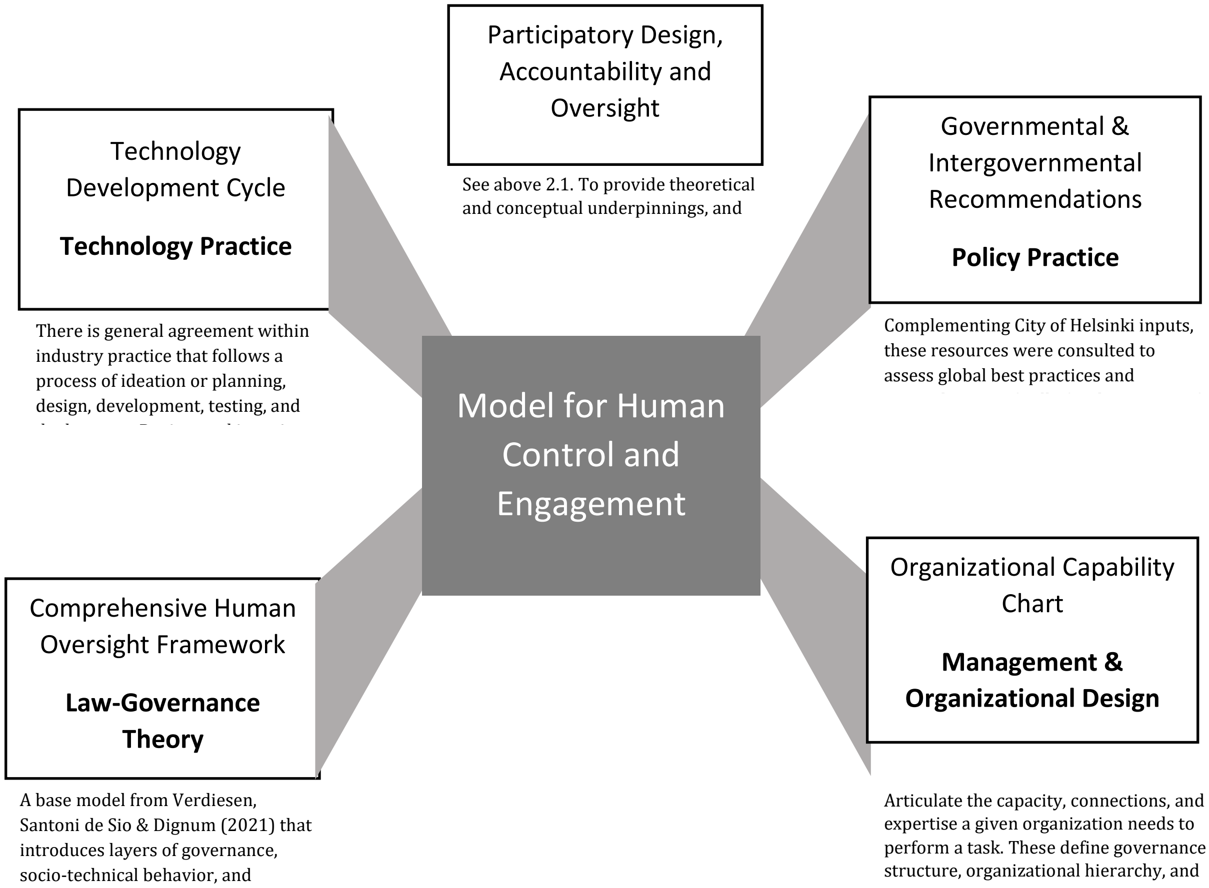


Figure 2: Model for Human Control and Engagement.

4.3 High Level Model for Human Oversight

The model consists of three layers that provide systems oversight. The main focus for human oversight in this model is on translation and escalation (Lassnigg, 2012) as knowledge articulation occurs through the practice of project governance (Brunet, 2019; Brunet & Aubry, 2018). The human oversight capacity includes a steering committee that translates ideas of accountable AI to governance and leadership priorities, includes built-in participatory design mechanisms that integrate external and expert feedback across multiple layers, and the ability to task specific working groups with translating relevant ideas of accountability to the technical layer. These groups work at different levels of abstraction within the system: on the one hand strategic and political and on the other operational and practical.

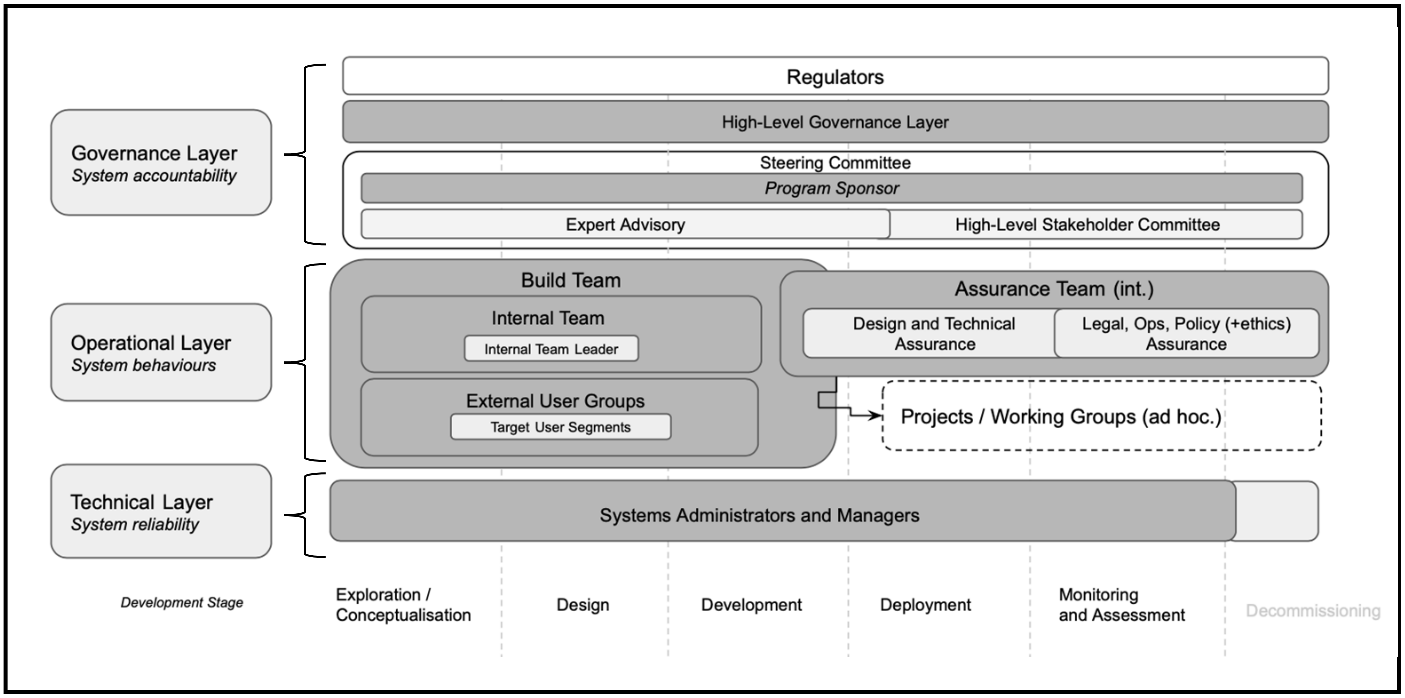


Figure 3: High Level Model for Human Oversight.

4.3.1 Model Features

The **Governance Layer** ensures high-level *system accountability*. Based on regulatory input, **Directors and Senior Leaders** set the vision, strategy and priority for oversight and accountability across the organization. A **Steering Committee** is also proposed to set high level priorities to task specific **Projects / Working Groups** as needed, on specific tasks, urgent matters, or in situations of immediate crisis. The committee and groups also play a translation role.

<u>Key responsibilities:</u>

- create governance structures within the organization,
- establish clear lines of accountability and reporting,
- enable high-level expert advisory and stakeholder feedback
- task Projects / Working groups on matters of strategic importance and priority

An **Operational layer** works to manage and tune *system behaviors*. Two persistent functional groups are envisioned at this layer, along with a third *ad hoc* capacity for matters of strategic importance. First, a **Build Team** consists of internal departmental staff (designers, builders and engineers) working with external user groups (teachers, parents, students). This team works to prioritize and develop new system features in the technology development lifecycle across exploration/conceptualization, design, and development phases.

Second, an Assurance Team will also overlap in terms of activities across some of these phases. The Assurance team holds capabilities for (a) design and technical assurance and (b) legal, operations, and policy oversight capability. The Assurance team plays a role in the technology development lifecycle through the effective deployment, monitoring and assessment of the system, and is responsible for initiating any steps for decommissioning features or systems.

Finally, the operational layer is supported by an ad hoc organizational capability for Projects / Working Groups with individuals who may be drawn from across the organization to execute on priorities of the Steering Committee.

Key responsibilities:

- define, design, and understand technology features and opportunities,
- manage technology risks,
- enable feedback channels and loops with key user groups,
- escalate governance errors or issues.

The last **technical layer** is responsible for maintaining *system reliability*. This layer includes systems administrators and managers. This functional layer works to configure operating systems and associated server hardware, software, databases and networks. In addition to daily upkeep, actors within this layer work to diagnose and troubleshoot technical network issues and provide technical support.

Key responsibilities:

- maintain system trustworthiness and availability,
- safeguard system security and reliability
- ensure technical robustness,
- resolve technical errors,
- report on system anomalies.

5 Translational Models

5.1 A model for operationalization of high-level AI ethics principles

Ethical AI has been in the center of the policy and regulatory debate on how to maximize the benefits and minimize the societal risks of the technology in the EU. Multiple ethical guidelines and best practices have been developed by different stakeholders, including governments, industry, civil society organizations, etc. Existing literature (Fjeld, Achten, Hilligoss, Nagy, & Srikumar, 2020; Zhou et al., 2020; Fukuda-Parr & Gibbons, 2021) has identified common principles and standards present in some of the recent AI ethical guidelines initiatives, which indicate that consensus on ethical governance frameworks for AI has increasingly become an important part of the development, use, and implementation of the technology.

Notwithstanding these debates, one of the biggest challenges of ensuring effective ethical AI in Europe is the operationalization of high-level ethical principles into actionable and practical measures. This challenge reflects different gaps in various levels of ethical AI governance which need to be addressed in order to guarantee that all actors involved in the multifaceted and highly complex problem of aligning technological innovation with safe, human centric, and ethical AI can effectively contribute to that goal.

The first gap identified in this phenomenon is between theory and practice, in other words, the gap between widely discussed ethical principles of human-centered AI (HCAI) and practical steps for effective governance (Shneiderman, 2020). Consequently, this indicates a second gap between policymakers and the AI industry, where policymakers have achieved consensus on the ethical principles they intend to prioritize, while companies focus on currently regulated issues rather than the ethical issues raised by AI (Miailhe et al., 2020).

Even though multidisciplinary research on this topic has substantially increased, there are still vocabulary and knowledge gaps between lawyers, legal scientists, computer scientists, and data scientists. Consequently, these gaps reflect the urgent need for more collaborations between the different fields involved in research and education on trustworthy AI. Finally, the last gap refers to the democratic and participatory challenges of stakeholders' participation in designing, developing, and deploying AI systems.

5.1.1 AI Ethics Systemic Translational Matrix

Considering the above-mentioned gaps and challenges, this report recommends a practical tool for translating high-level ethical principles into suggested mechanisms for technical builds, design, and participatory processes. The *AI Ethics systemic translational matrix for AI and Learning Analytics at Vocational Education and Training (VET) in Helsinki* (Appendix I) is a theoretical model for multilevel translation of the EC High-Level Expert Group on AI's (2019) seven key requirements which AI systems should meet in order to be deemed trustworthy.

The model is a matrix composed of seven rows, one for each ethical requirement (Human agency and oversight; Technical robustness and safety; Privacy and data governance; Transparency; Diversity, Non-discrimination, and fairness; Societal and environmental wellbeing, and Accountability), and seven columns. Each column represents one level of applied translation, using a top-down approach, starting from the high-level ethical requirements, through the identification of the concrete societal risks and harms addressed, towards the technical interpretation and stakeholders' participation methods (See figure 4).

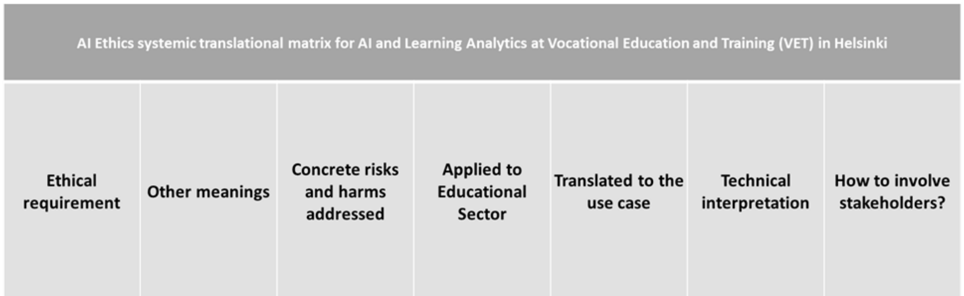


Figure 4: Translational Categories of the AI Ethics Systemic Translational Matrix for AI and Learning Analytics at Vocational Education and Training (VET) in Helsinki.

Two columns of the Matrix are flexible and should always be adapted according to the use case (column number four *"Applied to Educational Sector"* and column number five "*Translated to the use case*"*)*. This flexibility represents one of the most relevant features of the Matrix: it is a translational model that can be universally applied to any sector and to any AI use case.

Taking into account the gaps described in the previous section, the Matrix is designed based on the following purposes:

a. Bridge the gaps between theory and practice, applying systemic translation of the EC High-Level Expert Group on AI's ethical requirements to the concrete use case;

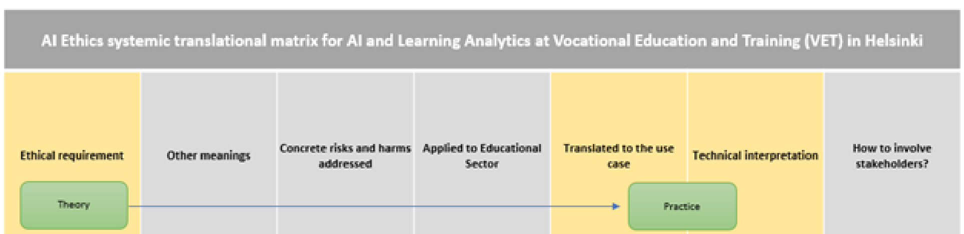


Figure 5: Visual demonstration of how the Matrix bridges the gaps between theory and practice.

b. Bridge the gaps between disciplines, by applying Human Rights-based approach and AI Ethics to Computer Sciences, Data Science, and Participatory Design methods;

Figure 6: Visual demonstration of how the Matrix bridges the gaps between disciplines.

c. Bridge the gaps between stakeholders by addressing the challenges of stakeholders' engagement and participation;

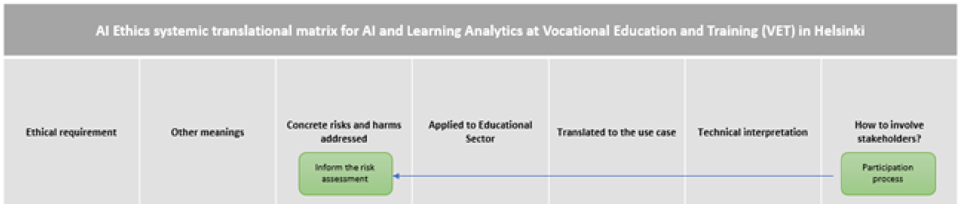


Figure 7: Visual demonstration of how the Matrix bridges the gaps between stakeholders.

d. Bridge the knowledge and vocabulary gaps between policymakers and technologists

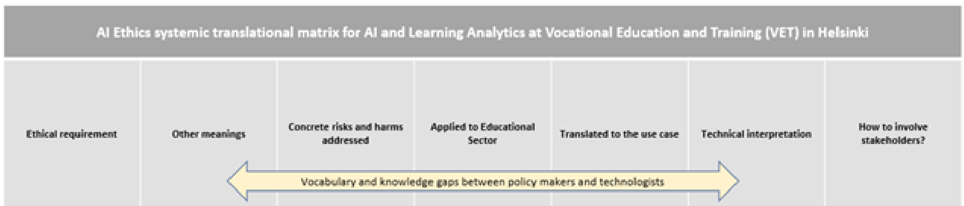


Figure 8: Visual demonstration of how the Matrix bridges the knowledge and vocabulary gaps.

5.1.2 How to use this tool

The complete Matrix can be found in **Appendix I**. The proposed translational model is composed of a combination of practical elements based on existing best practices, multidisciplinary literature, ethical and human rights-based frameworks, laws, and regulations. They are oriented to the design of human-centered AI systems that are reliable, safe, and trustworthy, which in turn bring benefits to organizations, individuals, and society.

This governance structure is a starting point. Newer approaches will be needed as technologies advance or when market forces and public opinion shape the products and services that become successful. It is not intended to be a static matrix, nor is it an exhaustive list of risks, tools, and requirements.

In order to make the best use of the Matrix, the Education Division of the City of Helsinki should continuously update and improve the model, which would be ideally informed by constant new contributions from stakeholders' participation (e.g., students expressing new concerns and teachers identifying new challenges while using the AI system); from the Technical AI Team, who would test and communicate new shortcomings; from the general public, who expresses their opinions and advocate for changes, etc. Additionally, upcoming policies, laws, and regulations in Finland and in the EU will present new legal or ethical requirements. Finally, the research community plays an important role when identifying new problems and proposing new theories, which should also inform this model.

All these actors and elements will constantly inform the systemic translational matrix in a reflexive and critical way; therefore, changes will be sustainably made towards the ultimate goal of minimizing the risks and maximizing the benefits of AI.

5.2 Translational model for Human Oversight measures

Article 14, paragraph 4 of the proposed AI Act states that “[T]he human oversight measures shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances”. All the five measures described in the Proposal are interpreted below in three different layers: technical layer, socio-technical layer, and governance layer.

The three layers of the translational model were informed by Verdiesen, Santoni de Sio, and Dignum (2021)’s Comprehensive Human Oversight Framework, which aims to operationalize the concept of Human Oversight in a comprehensive approach and to provide concrete recommendations for an oversight process.

The human oversight measures described in Article 14, paragraph 4 of the proposed AI Act were selected as the conceptual foundation for the translational model because they correspond to one of the existing understandings on what human oversight should look like. Furthermore, the EC Proposed AI Act was informed by the EC High-Level Expert Group on AI (2019)’s seven ethical requirements of Trustworthy AI; therefore, this theoretical approach aligns with our goal of consistently applying an ethical and rights based-approach to AI.

The human oversight measures are translated as follows:

Article 14, paragraph 4, Proposed AI Act: *“The Human oversight measures shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances”.*

(a) fully understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible;

- Technical layer:

Engineers of the system will be able to accurately describe what the AI system can and cannot do by describing the bounds of the data sets and its limitations. Engineers will build in test-cases and code that can detect unexpected outputs and failures of the system. Engineers will create a reporting system for assigning unexpected behavior tickets to engineers to fix.

During the testing phases, engineers should also test with biased inputs to measure the threshold of the accuracy in the system’s results and outputs. This can also assist developers in identifying what a misuse of the system would look like. This can help to recheck the data elements classification from which the results are derived.

- Socio-technical layer:

Educators and other professional educational staff that interact with the AI system should receive training on expected outputs, how they may detect unexpected behavior of the system, who to report unexpected outcomes to, and what non-AI systems to fall back on should the system behave in unexpected ways.

- Governance layer:

The measure described as "fully understand the capacities and limitations of the high-risk AI system" could be implemented by using counterfactual explanations, which specify what circumstances would need to change to achieve a more desirable decision, in contrast to explanations that involve an attempt to outline the logic of algorithms. Counterfactual explanations attempt to address the human interpretability issues inherent in machine learning algorithms. Counterfactual explanations do not require individuals to understand any algorithms in order to extract a meaningful explanation. They are easy to understand and practically useful as they provide the circumstances that need to change to achieve a more desirable decision. (Gacutan & Selvadurai, 2020).

(b) remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system ('automation bias'), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;

- Technical layer:

Engineers should be made aware that the eventual goal of the system is not assessment, but identification of at-risk students for outside services to assist them in their education. It should be built into the system notifications for when someone is interacting with AI or when they are viewing a suggestion or claim made by an AI and not a human.

When providing the recommendations for at-risk students, the system should be capable of explaining in simple language what are the factors that the AI considered in order to make the recommendation as appropriate to the circumstances and existing interpretability capabilities (Watson, 2020). Understandability of the explanations provided is another critical factor that engineers should be aware of when developing systems for sensitive user groups, such as students.

- Socio-technical layer:

Educators should be encouraged to push back on and question the decisions of the AI system. For instance, if an instructor feels that a student would have a better school experience with additional support, but the AI system has not identified that student for support, the instructor should feel free to offer that student support services.

Educators and managers should come to an agreement about lines of accountability for decisions made or informed by algorithms to overcome issues related to blame avoidance (Hood, 2007), or deference to algorithmic recommendations (Zuboff, 1988).

- Governance layer:

This measure requires that human oversight is ensured in such a way as to enable the person assigned that task to be aware of the potential of 'automation bias'. Automation bias occurs when an operator relies solely on automated recommendations without searching for disconfirming evidence (Davis et al., 2020).

The explicit recognition of this problem is valuable in itself. Yet, combatting it more effectively might necessitate additional safeguards, for instance by requiring the Education Division of the City of Helsinki to communicate how other available information or alternative outcomes were considered in reaching a decision. (Fink, 2021).

(c) be able to correctly interpret the high-risk AI system's output, taking into account in particular the characteristics of the system and the interpretation tools and methods available;

- Technical layer:

In the design of the system, the outputs should be immediately understandable even to those with no technical background. For instance, converting psychometric outputs to plain, easy to understand language. In the design of the dashboard, it should be made clear what data sources are being used in an AI system, the limitations of that data, and how an output is constructed from those inputs.

The use of jargon and technical terminologies should be avoided in the system's explanation to make it understandable for any audience (Mittelstadt et al., 2019).

- Socio-technical layer:

Educators should be trained on expected outputs for on-track and at-risk students, and how those outputs translate into access to real world support systems for graduation success. For instance, in the case of different levels of at-risk identifications, what supports would a teacher be able to call upon for different kinds of students?

- Governance layer:

The right to explanation envisaged in the European Union's 2018 General Data Protection Regulation (GDPR) allows an individual to seek 'meaningful information' about the 'logic' involved in making a decision, in circumstances where the decision was made solely using automated technologies and the decision produced legal effects concerning the individual or significantly affected them. Although the legal status of this right to explanation has been the subject of considerable debate, it is an important development towards human oversight, interpretability and explainability.

Explanations of specific algorithmic decisions should allow the justification of a black-box model or decision to be debated and contested. Further, meaningful, critical dialogue must be achieved between user, developer, and model by ensuring explanations are contrastive, selective, and social. (Mittelstadt et al., 2019).

Participation—including the related requirement of information transparency—and accountability are inter-related principles that build on each other in the practice of human rights; it is only when people have the information and can participate in decisions that the designers and users of AI design, development and deployment can be held to account. (Fukuda-Parr & Gibbons, 2021).

(d) be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system;

- Technical layer:

Engineers should design a system that allows educators access to non-AI decision making data, so that users can fall back on their own data analysis should the system behave in unexpected ways. In the pre-implementation stage where engineers are training models, they should have access to the governance layer to explain why they believe a model may not work or cause unintended harm.

- Socio-technical layer:

Educators should still be able to make their own decisions about what students need and what additional supports would be best for their academic success. Educators should have access to engineers to express their concerns about when to disregard or override the suggestions of the AI system.

Questions about whether students, parents, and public sector managers should also be able to make decisions about when not to use high-risk AI systems, for example around automated or semi-automated support, or dashboards, will also need to be addressed.

- Governance layer:

The human-in-command (HIC) approach refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal, and ethical impact) and the ability to decide when and how to use the system in any particular situation. This can include the decision not to use an AI system in a particular situation, to establish levels of human discretion during the use of the system, or to ensure the ability to override a decision made by a system.

Human oversight needs to be meaningful in the sense that the overseer should have the authority and competence to change the decision (Brkan, 2019).

This human oversight measure observes the fundamental right to an effective remedy (Art. 47 EU Charter of Fundamental Rights) and the GDRP, Art. 22, par. 3 “The data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.”

(e) be able to intervene on the operation of the high-risk AI system or interrupt the system through a “stop” button or a similar procedure.

- Technical layer:

The system should be designed such that the AI functionality can be completely pulled from AI-HOKS and not interrupt the student data stream. While the AI system is offline for any reason, educators and students should still have access to their VET data. This can be designed by integrating the AI system in a modular way into AI-HOKS to not disrupt the service.

During the occurrence of any adverse operation, the AI recommendations or suggestions should be turned off. Users should still have access to the service where they can interact such as, viewing their grades, past/current/upcoming courses, etc. Whereas the recommendation AI should be deactivated and directed to the testing team for error identification.

- Socio-technical layer:

In the event of a system stop, educators should be trained on how to access the data in other sources or databases should AI-HOKS go down. If a user believes that the system is acting in a way causing widespread harm, then they should have access to an escalating system that alerts engineers and governance stakeholders.

When the recommendations and suggestions provided by AI are halted because of anomalies, students should still be able to see their grades, course lists, and other basic course/personal information without interruption. AI used here should only enable the students with additional support rather than taking away the basic access to the service.

- Governance Layer:

This measure highlights the need for and importance of human autonomy when applying oversight to AI systems. The overseer should have the capacity, autonomy, and power to be able to intervene and stop the AI system.

It is important to address the existing power relations when human oversight is being deployed. If individuals to whom human oversight is assigned do not have enough autonomy to effectively intervene in AI operations, the purpose and efficacy of oversight measures will be compromised, and their impact will be undermined.

6 Practical Tool: Public Engagement and Education Site

6.1 About this tool

Helsinki has existing digital initiatives that address participation and oversight for technological advancements (e.g., websites such as the City of Helsinki AI Register or Design Helsinki, and tools such as the one developed by Saidot). Building on Helsinki's digital infrastructure, we here aim to display how to further develop future participatory websites, i.e. how to add features and functionality so that participation and human oversight capacities can be merged in one single platform. We call this proposed platform / website *EdTech Helsinki*. Firstly, the website is a way for residents to view ongoing initiatives in the field of education technology, the participatory process, see how they can get involved, identify areas where they might be interested in participating, and offer suggestions where they see gaps or misunderstandings. Secondly, it serves companies who aim to develop education technology by providing guidance on how to set up participatory designs and by allowing them to obtain feedback from citizens and oversight bodies. Thirdly, it helps oversight bodies to monitor and evaluate the progress of ongoing education technology initiatives. Ultimately, the aim is to enhance participation and support human oversight.

More information can be found in the detailed user guide in **Appendix II**: User guide for EdTech Helsinki platform.

7 Appendixes

7.1 Appendix I: AI Ethics Systemic Translational Matrix

AI Ethics Systemic Translational Matrix for AI and Learning Analytics at Vocational Education and Training (VET) in Helsinki

Ethical requirement (High-Level Expert Group on AI, 2020)	**Other meanings** (Fjeld, Achten, Hilligoss, Nagy, and Srikumar, 2020)	**Concrete risks and harms addressed** (risk-based approach and human rights-based approach)	**Applied to Educational Sector** (Vincent-Lancrin, S. and R. van der Vliesm, 2020; Slade, Sharon & Tait, Alan, 2019; Miao, Holmes, Huang, Zhang, 2021)	Translated to the use case	**Technical interpretation** (European Commission. Directorate General for Communications Networks, Content and Technology., 2020, Weyns, 2020, Gotterbarn et al., 1999, Thomson & Schmoldt, 2001)	**How to involve stakeholders?**
1. Human Agency and Oversight	"human review of automated decision," "ability to opt out of automated decision," and "human control of technology."	Human out-of-the-loop. Decisions are not explainable. There is no obvious accountable party. Automatic decisions cannot be contested by the person they impact, nor can they be modified in exceptional circumstances by public administrators.	Student agency and responsibility (Slade & Tait, 2019).	Although it is clear that there is an asymmetrical power-relationship between institutions and students, proactive engagement at least seeks to treat students as equal participants in the uses of their data. In this way, students can be more actively involved in helping the institution to design and shape interventions that will support them. (Slade & Tait, 2019).	Users have control about how their data enters the system and understands the lifecycle of their data. Additionally, an understanding of how AI processes their data. It is clear to the user whether they are interacting with a human or AI. How is human oversight implemented into the design?	Use participatory tools that could elicit information about whether users feel reductions in their agency or situations where the technology exhibits undue control. Ice breaker exercises, Contextualizing with scenarios (Scenario-based approach), embodied participatory methods (e.g., walkthrough or media go-along) (Malinverni et al., 2019; Light et al., 2018; Jørgensen, 2016)
2. Technical Robustness, Accuracy and Safety	"safety," "security," "security by design," and "predictability".	Risks of a negative impact because of unreliable or low quality decisions. Risks of Cyber-attacks (e.g., ransomware, denial of service, data breach). Accuracy amounts to understanding performance, identifying the sources of error and the limitations of a solution and considering the quality and reliability of the decisions, as well as their direct societal impact. (Unceta, Nin and Pujol, 2020)	Students seem to be quite positive about the possibilities of learning analytics but are also concerned about the safety and usage of their personal data. (Nevaranta, Lempinen, & Kaila, 2020).	Inaccurate data could lead to inappropriate evaluations of progress or recommended supports. Security breaches could release personal information that could lead to harms (such as stigmatization or discrimination).	Relevant safety measures are built into the codebase and backend. There are protocols to deal with any potential breaches. Cyber-security measures have been taken in accordance with EU law. The technology is accurate and there are mechanics to monitor or improve accuracy.	Use participatory tools that could elicit the most relevant information about where risks of data inaccuracy or data security could lead to the greatest harms. Interviews, focus group discussion, Online surveys, Observation
3. Privacy and Data Governance	"privacy by design," "consent," "control over the use of data," "ability to restrict data processing," "right to rectification," "right to erasure."	Risks to the **right to private life** (Art. 7) and the **right to the protection of personal data** (Art. 8, EU Charter of Fundamental Rights). This dimension accounts for risks in three forms: **reidentification risk, data linkage risk**, and **sensible attribute inference risk**. Reidentification considers the probability of identifying an individual in the training set. Data linkage concerns the probability of being able of linking/joining two different datasets. Sensible data inference concerns the problem of using a ML system to infer protected information. This risk involves the leakage of sensible information through other attributes (Unceta, Nin and Pujol, 2020).	The collection and storage of data create new risks for privacy of students. Beyond the "Big Brother" fears that are common to all sectors of society, additional concerns related to privacy and AI in education usually are at least twofold. Families are concerned that education institutions or even employers may use "old" data to make decisions, which raises the question of how long and which data could be stored and retrieved to make some decisions. A second question relates to the possible use of the data for commercial purposes in a sphere where commercial interests are often excluded. (Vincent-Lancrin and Van der Vlies, 2020).	Students should have some input to determine which data can be collected, how that data can be used, who is able to access it, and for what purposes (Prinsloo & Slade, 2017). CoH should grant students the ability to correct and/or add context to their raw data, and to review and make a case for choices which appear to be limited as a result of a learning analytics application.	Users can pull their data from the system at any time, and there are protocols in place to implement this. Oversight mechanisms for data processing (including limiting access to qualified personnel, mechanisms for logging data access and making modifications). Measures to achieve privacy-by design and default (e.g. encryption, pseudonymisation, aggregation, anonymisation). Establish mechanisms that allow flagging issues related to privacy concerning the AI system.	Use participatory tools that could elicit the most relevant information about individual privacy preferences (both expressed and enacted). Identifying and prioritizing benefits and privacy risks of using LA through **participatory design**, by applying **Nominal Group Technique (NGT) with students**, exploring their perceptions of privacy protection and the use their data. In the case of privacy risks, understanding the perspective of the data subjects (eg, students) is critical because the privacy-enhancing design options are predominantly for their benefit. Moreover, because individuals' privacy preferences often conflict with their actual behaviors (ie, the privacy paradox), understanding which **privacy-enhancing design options** are most critical cannot be accomplished without proper engagement. This engagement not only leads to better software, but also participatory software design can importantly increase transparency and trust in AI. (Giannouchos et al., 2021).
4. Transparency	"transparency," "explainability," "open source data and algorithms," "open government procurement," "right to information," "notification when interacting with an AI," "notification when AI makes a decision about an individual," and "regular reporting."	A lack of transparency about how tools work could lead to a decline in public trust (Janssen et al., 2020), or to system avoidance behaviours (Brayne, 2014).	Transparency can be considered as one of the most crucial factors regarding the acceptance of learning analytics systems: This involves disclosing information about the collected data, its purpose, the underlying algorithms, the people who receive access to the data, and the analyses derived from them, as well as the amount of time the data will be stored and its degree of de-identification (Pardo and Siemens, 2014).	If teachers do not know if or how the tool will be used to manage their performance, they may resist its use (Jakobsen et al., 2018; Mumtaz, 2000).	Building warnings for the user when they are interacting with AI, making data open source where possible, making code open source where possible, transparent data cleaning processes. The decisions the AI makes are understanble and traceable in such a way that the average user can understand how their data is processed. The purpose of the AI is communicated clearly to users.	Use participatory tools that could elicit information about whether users feel there is sufficient transparency or to help identify where they would like to see more information available. Storyline workshops, Contextulizing with scenarios (Scenario-based approach), web-based tools to facilitate participation and engagement (Viale Pereira et al., 2017).
5. Diversity, Non-discrimination and Fairness	"non-discrimination and the prevention of bias," "representative and high-quality data," "fairness," "equality," "inclusiveness in impact," and "inclusiveness in design."	Risks to the **right to equality and non discrimination** (Article 21, EU Charter of Fundamental Rights). Risk of perpetuating and **amplifying existing societal bias** and **discriminations** against certain collectives or minority groups. Risk of applying a false computer neutrality to an algorithmic decision based on biased datasets. This dimension ensures that algorithmic decisions do not display an unjust or biased behavior with respect to sensitive factors such as gender, race or religion. AI technologies have deep reach and can transform political, economic and social institutions of the 21st century. Used by and serving the interests of the powerful, whether it is the state or a corporate actor, artificial intelligence's design, development and deployment (AI-DDD) reinforces power structures and can enable oppression of the vulnerable rather than their protection and empowerment. (Fukuda-Parr & Gibbons, 2021). There is also a risk of lack of information sharing that leads to harm (inability to audit for bias in machine learning) (Benthall & Haynes, 2019).	The models used to analyse, interpret and communicate learning analytics to stakeholders (support staff, advisers, faculties, students) should be sound, free from algorithmic bias; transparent where possible and clearly understood by the end users (Slade & Tait, 2019).	Early warning systems powered by AI will typically profile students and identify who is at risk of dropping out. If their effectiveness in identifying the right students is too limited, even if they do no more harm than the lack of a system, they are not fully trustworthy and need improvement through further research and development. Another possibility is that they are accurate but misused. Identifying who is at risk of dropping out matters only if a good (human) intervention to support the students and address that risk is implemented. (Vincent-Lancrin and Van der Vlies, 2020).	Datasets that represent the field of possible users in training, testing, and validation sets. Including minority voices in model training. Ways to mitigate bias in training by educating developers on possible bias in their models. Using Universial Deisgn for Learning to allow all kinds of users to interact with the system. Stakeholders are consulted on the design of the technology.	Use participatory tools that could elicit information about how different groups experience discrimination and use those insights to explore what fairness and diversity would look like in the context of the AI initiative. Is your definition of fairness commonly used and implemented in any phase of the process of setting up the AI system? Did you consider other definitions of fairness before choosing this one? Did you consult with the impacted communities about the correct definition of fairness, i.e. representatives of persons with disabilities?
6. Societal and Environmental Well-being	"environmental responsibility"; "Impact on Work and Skills"; "Impact on Society at large or Democracy."	Ubiquitous exposure to social AI systems in all areas of our lives (be it in education, work, care or entertainment) may **alter our conception of social agency**, or **negatively impact our social relationships** and attachment. While AI systems can be used to enhance social skills, they can equally contribute to their deterioration. This could equally affect peoples' physical and mental well-being. (EC High-Level Expert Group on AI, 2020). There is also a risk of lack of information sharing that leads to harm (inability to identify harmful patterns that could be prevented through service interventions) (Doll, 1974; Parkin & Paul 2011).	Governments must work with stakeholders to shape AI in education to help prepare for the transformation of the world of work and society. (Vincent-Lancrin and Van der Vlies, 2020).	There may be some educational information related to progress and social connection that may lead a teacher or other education professional to recommend health or social service interventions. It may be necessary to use and share data for these purposes.	Developer considers technology's environmental or societal context, such as energy use and carbon emissions, and its impact on the humans who will use the technology. Technology that does not have negative impacts on democracy (such as by amplifying fake news).	Use participatory tools that could elicit information about under what circumstances education data and AI outputs could be used for secondary purposes to support meaningful service interventions that could improve wellbeing. Citizen juries (Parkin & Paul, 2011), embodied participatory methods (e.g., walkthrough or media go-along) (Malinverni et al., 2019; Light et al., 2018; Jørgensen, 2016)
7. Accountability	"verifiability and replicability," "impact assessments," "evaluation and auditing requirements," "creation of a monitoring body," "ability to appeal," "remedy for automated decision," "liability and legal responsibility," and "accountability per se."	Risks related to **interpretability** and **explainability**. The first refers to a measure of the white-boxiness of a model. The second seeks the verbalization of algorithmic decisions at different levels of abstraction, corresponding to the different knowledge and needs of stakeholders, regulators and end-users. It accounts for the risks of ensuring that algorithmic decisions can be contested and reasoned upon. (EU Charter of Fundamental Rights, Article 47, **Right to an effective remedy** and GDPR, Art. 13-15, **Rright to explanation**). The the idea of explainability often transcends the ML models themselves to include not only the technical but also the human dimension (Unceta, Nin and Pujol, 2020)	Explainability (sometimes called interpretability) that goes beyond satisfying students' desire to understand the application and legal requirements to provide explanations (GDPR, Art. 13-15, Rright to explanation). Explainability helps designers enhance correctness, identify improvements in training data, account for changing realities, support students in taking control, and increase user acceptance. (Weld and Bansal, 2019).	A student may wish to understand why they are receiving a certain rating on their progress or a particular support. They may wish to remedy a perceived data error.	Regular assessment of the tools built. Outside parties that can evaluate the tool for biases and effectiveness. Programmers and designers who understand the legal obligations of AI. Reliable human-centered AI systems are produced by applying sound technical practices to software engineering teams. These technical practices clarify human responsibility, such as audit trails for accurate records of who did what and when, and histories of who conducted design, coding, testing, and revisions. (Shneiderman, 2020).	Use participatory tools that could elicit information about how stakeholders would like to see the lines of accountability for the operation and recommendations of the AI tool. Contextualizing with scenarios (Scenario-based approach), sentiment analysis (Ingrams, 2020).

7.2 Appendix II: User guide for EdTech Helsinki platform

EdTech Helsinki

User guide for implementing the proposed platform

Table of content

1 Introduction

This user guide elaborates on the proposed platform / website called *EdTech Helsinki*. In particular, this user guide is structured as follows: (1) introduction, (2) platform / website front-end and (3) summary.

For each section, we will here provide a short text that summarizes key information. In addition, we will include a box that states "limitations / things to consider" as well as "relevant back-end material" where applicable. The information provided in "limitations / things to consider" should draw attention to issues that – due to time constraints – couldn't be further developed during this project, yet still are important to be attended to in the future. The links stated in "relevant back-end material" refers to the sections within the main body of the "Summer 2021 AI Policy Research Clinic – Policy Paper" where theoretical grounding of this proposed platform can be retrieved. Thus, the research report – to some extent – represents the back end of the proposed platform / website.

The focus of this user guide is the visualization and elaboration of key features and how to turn theory stated in this policy report into practice. By providing all these types of information, we hope to warrant easy implementation of the proposed platform in the future.

1.2 Access & users of the platform / website

The platform is addressed to three different types of users*:

(1) EdTech launchers – These include any individuals who are currently developing or aim to develop an education technology (here called 'initiative') that will be deployed in Helsinki, regardless of whether these individuals are working in private or public organizations
(2) Community – This includes anyone who is interested in using the platform / website
(3) Oversight team members – This includes all members that are determined by the City of Helsinki

The platform could be integrated into the existing Helsinki AI Register infrastructure (https://ai.hel.fi/en/ai-register/) or could be accessible freely and without any barriers, for example, on https://edtech.hel.fi/en/. Solely the interface for the EdTech launcher users as well as the oversight team members would require an additional login.

1.3 Aim of the platform / website

The main aim of the platform / website differs according to the type of user:

- EdTech launchers: Ad hoc feedback & assurance
  - Receiving guidance on how to set up a participatory toolkit, including sample steps, questions and guidelines; registering an initiative; receiving ad hoc and ex post feedback from community and oversight team members on necessary actions
- Community: Transparency, inclusion & participation

 - Receiving the ability to keep up to date with, engage in and shape the developments of ongoing education technologies by providing feedback through the system or by identifying opportunities to participate at various stages of the project lifecycle and expressing interest in participating
- Oversight team members: Transparency & simplification
  - Receiving the ability to keep track of all ongoing EdTech developments; being reminded once oversight activities are required; allows giving feedback in a simplified and structured manner

Furthermore, the overall aim of a technological solution – this proposed platform – is fivefold:

- Universal transparency: everyone can access all of the information at any time, all past and ongoing initiatives are registered / centralized on one website so that no information is lost (regardless of whether the initiative is in place, abandoned, or replaced by something new)
- Collective knowledge generation: since anybody using this platform can provide comments / suggestions that are then transferred to the hosts of the website, the platform itself, i.e. the suggestions it is making can improve and become more holistic over time
- Standardization: by providing prefabricated ways of communication, participation and evaluation, information about all initiatives will be submitted and available in the same structured manner
- Efficiency: by providing EdTech launchers, the community and the oversight team a platform that sketches out a fixed process (i.e. the technology life cycle) as well as necessary forms of communication, participation and evaluation, the users do not have to be occupied with thinking about how to set up a particular process but can deep dive into the actions that are defined for each stage
- Collaboration and learning: information, best practices, and lessons learned about the initiatives can be used by other departments in the City of Helsinki, or by other municipalities in Finland when trying to initiate or deploy their own AI initiative

Limitations / things to consider:	• General remark: Do not use the icons/ pictures displayed in these mockups for further distribution / commercial activities since they were not officially bought. The links to the source where a particular icon / picture was retrieved from is stated in the speaker notes of the corresponding slide in the mockups.
Relevant back-end material:	• *For more information, see “Figure 3: High Level Model for Stakeholder Engagement”: The EdTech launchers represent the technical layer and – partly

[redacted]	– the build team at the operational layer, the community represents – partly – the build team at operational layer, the oversight team members represent the assurance team at the operational layer as well as the governance layer.

2 Platform / website front-end

In this chapter, we will visualize what the proposed platform *EdTech Helsinki* could look like and provide brief descriptions for each screenshot.

2.1 Landing page

The landing page (Figure 9) of *EdTech Helsinki* will be similar to Design Helsinki.

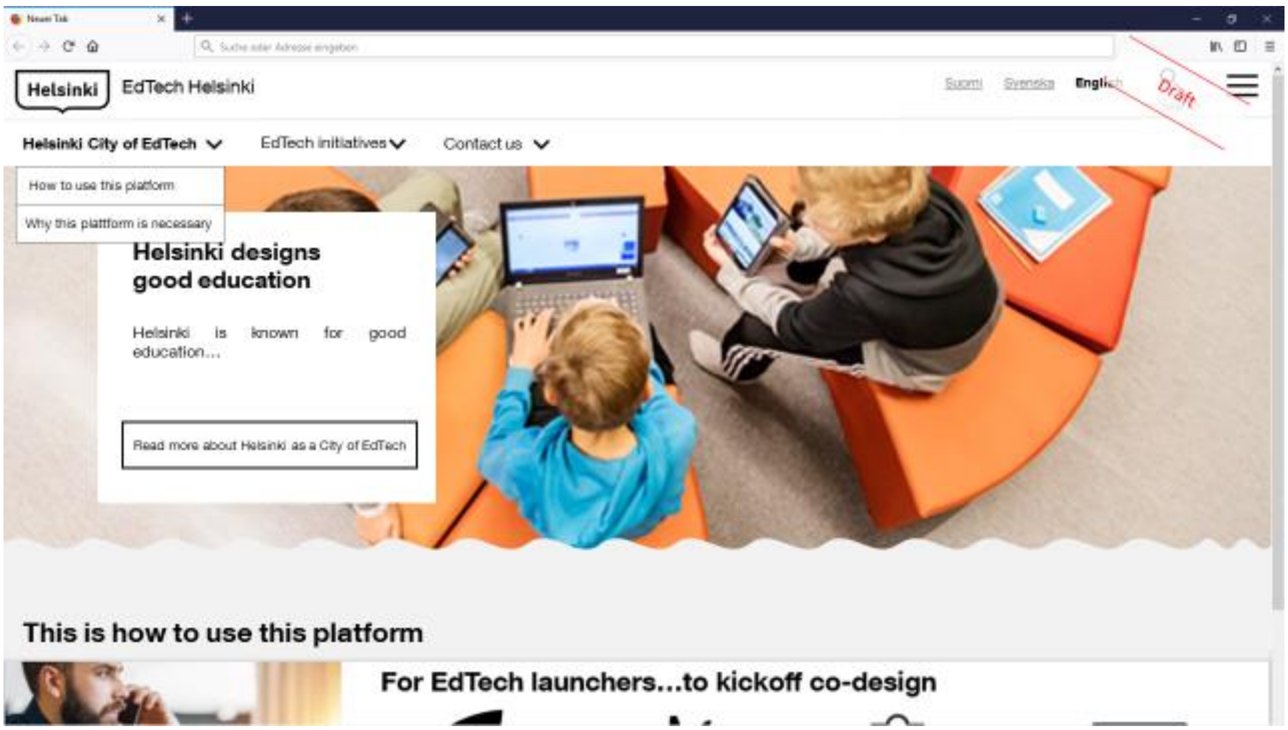


Figure 9: Landing page of open access platform

The landing page also displays information on how this platform can be used (Figure 10) and why this platform is necessary (Figure 11).

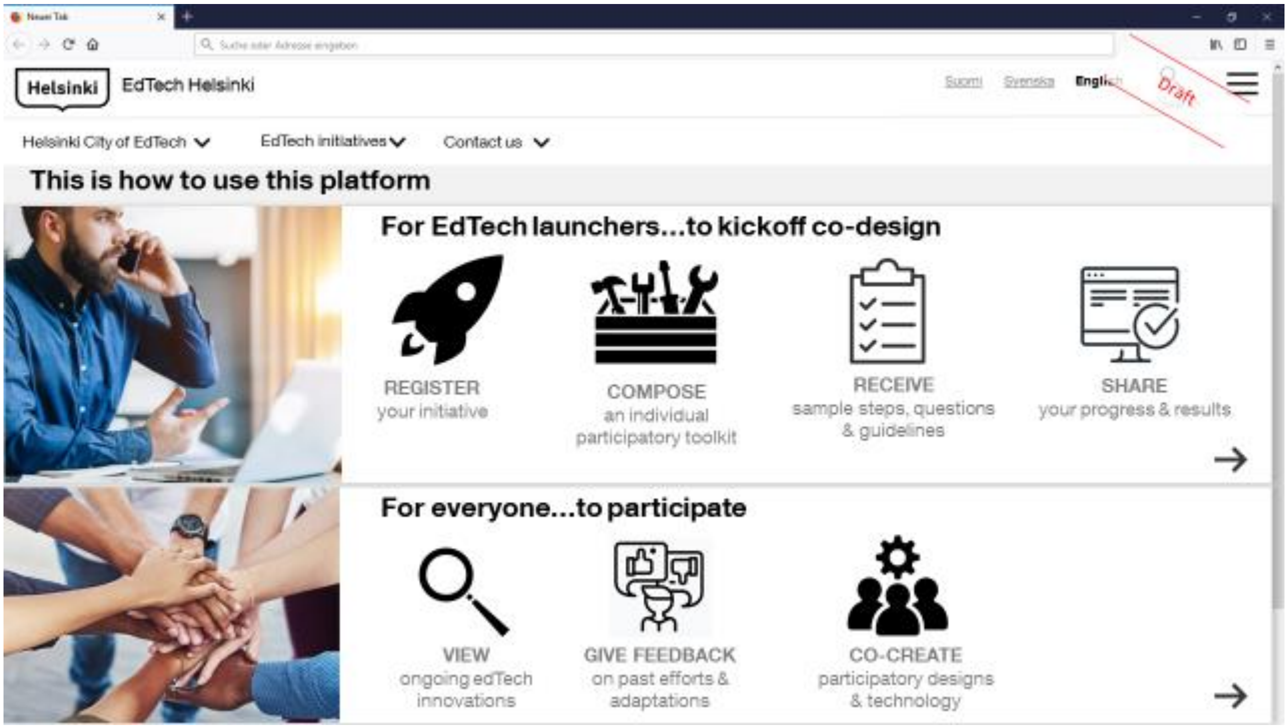


Figure 10: "This is how to use this platform"

When clicking on the arrow, users will be forwarded to the corresponding webpage. For the arrow of EdTech launchers, users will be forwarded to create their own initiative (Figure 12); for the arrow of everyone, users will be forwarded to the EdTech initiatives register (Figure 31).

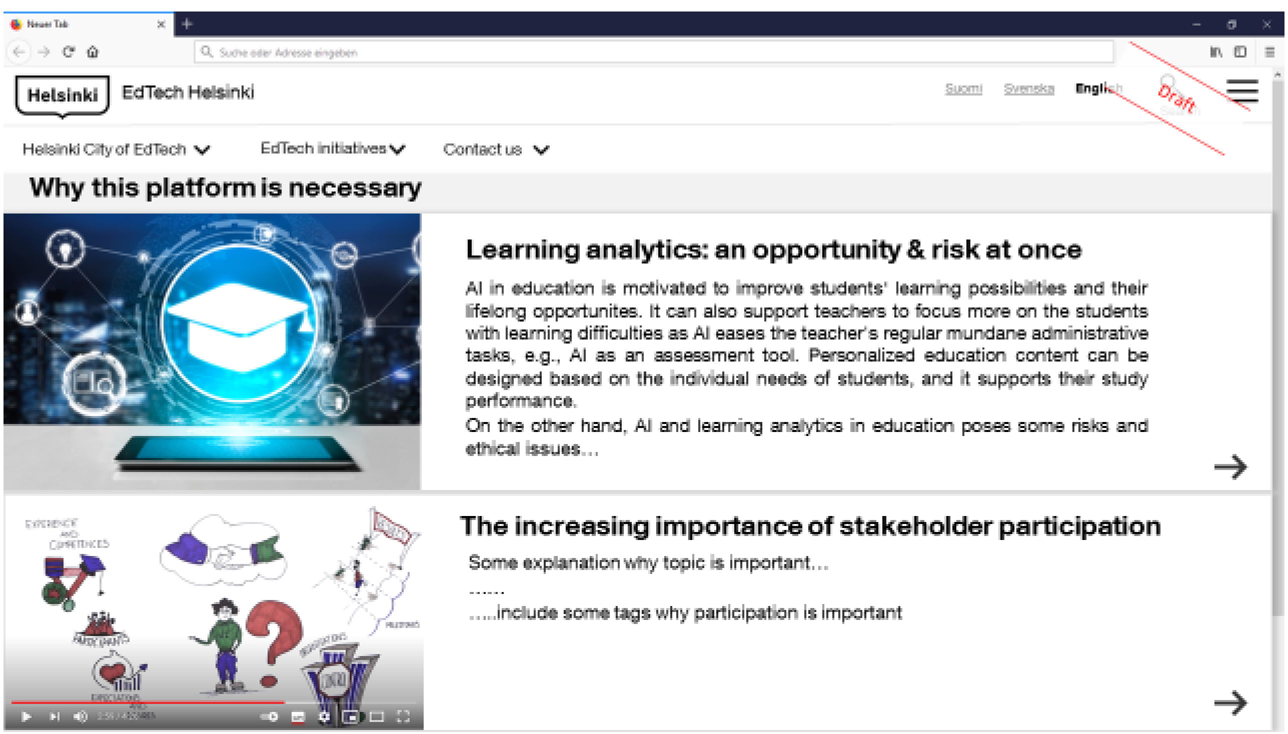


Figure 11: "Why this platform is necessary"

Information that will be displayed here is some key information on learning analytics* as well as the increasing importance of stakeholder participation** (Figure 11).

Relevant back-end material:	● *For more information, see "Learning Analytics", "Artificial Intelligence" and "Technology in Education". ● **For more information, see "Participatory Design" and "Key Stakeholders in Education".

2.2 Interface for EdTech launcher users

On the tab "EdTech initiatives", users can create a new / own initiative (Figure 12).

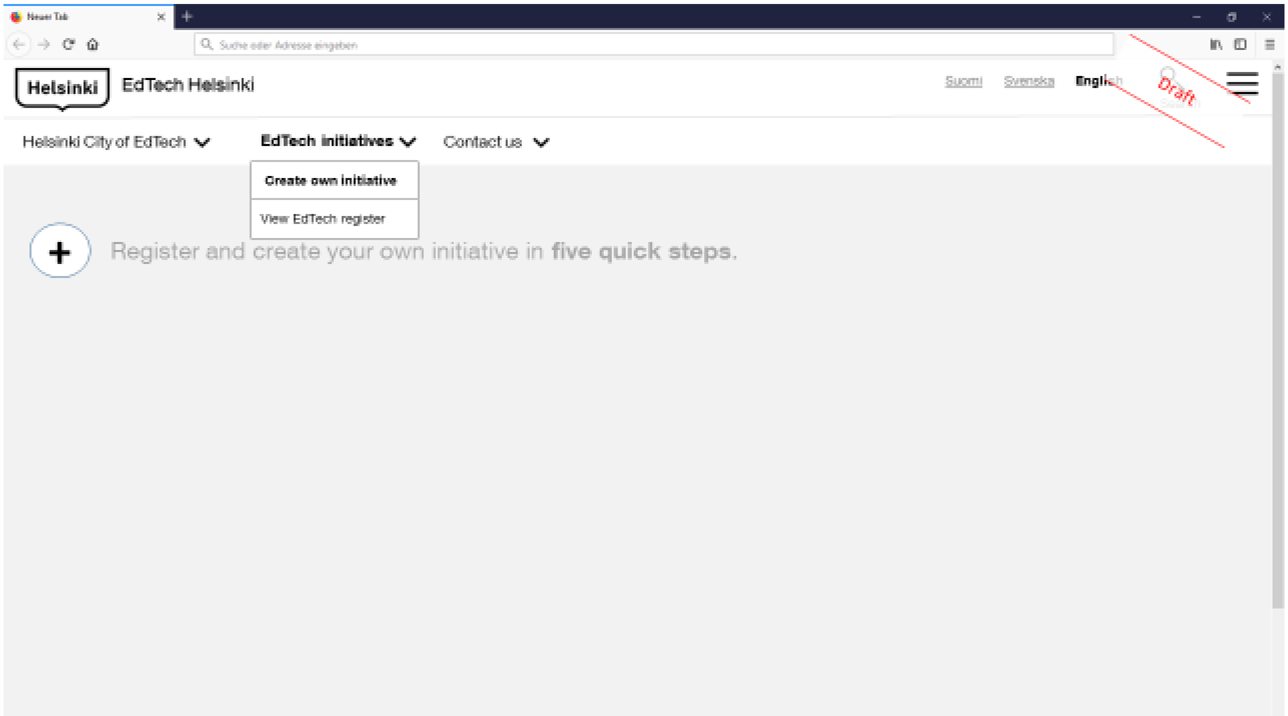


Figure 12: Creating a new initiative

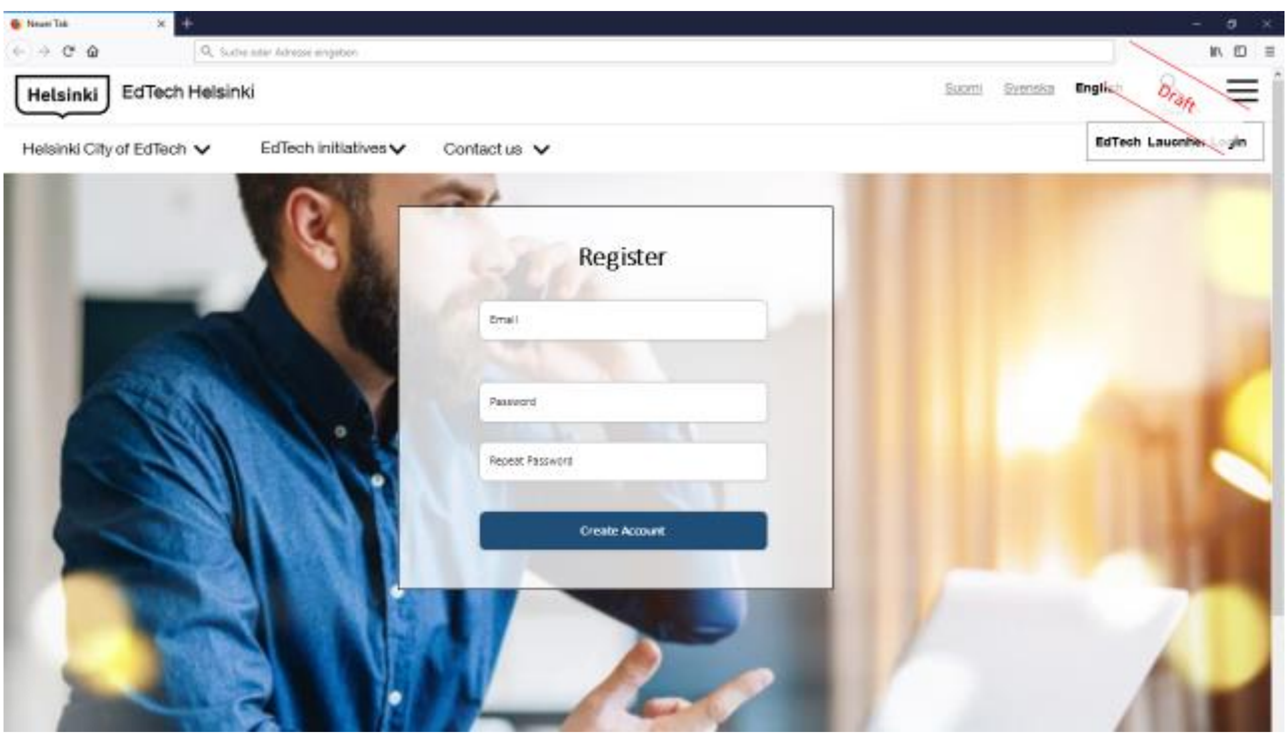


Figure 13: Account creation for EdTech launcher

After registering (Figure 13), the EdTech launcher can view its overview page (Figure 14) and start the creation process of a new initiative that takes five steps: (1) Set up initiative, (2) Select current stage, (3) Compose participatory toolkit, (4) Adjust participatory toolkit and (5) Finalize participatory toolkit. These steps will be elaborated in more detail in the following.

2.2.1 Setting up initiatives & creating participatory toolkits

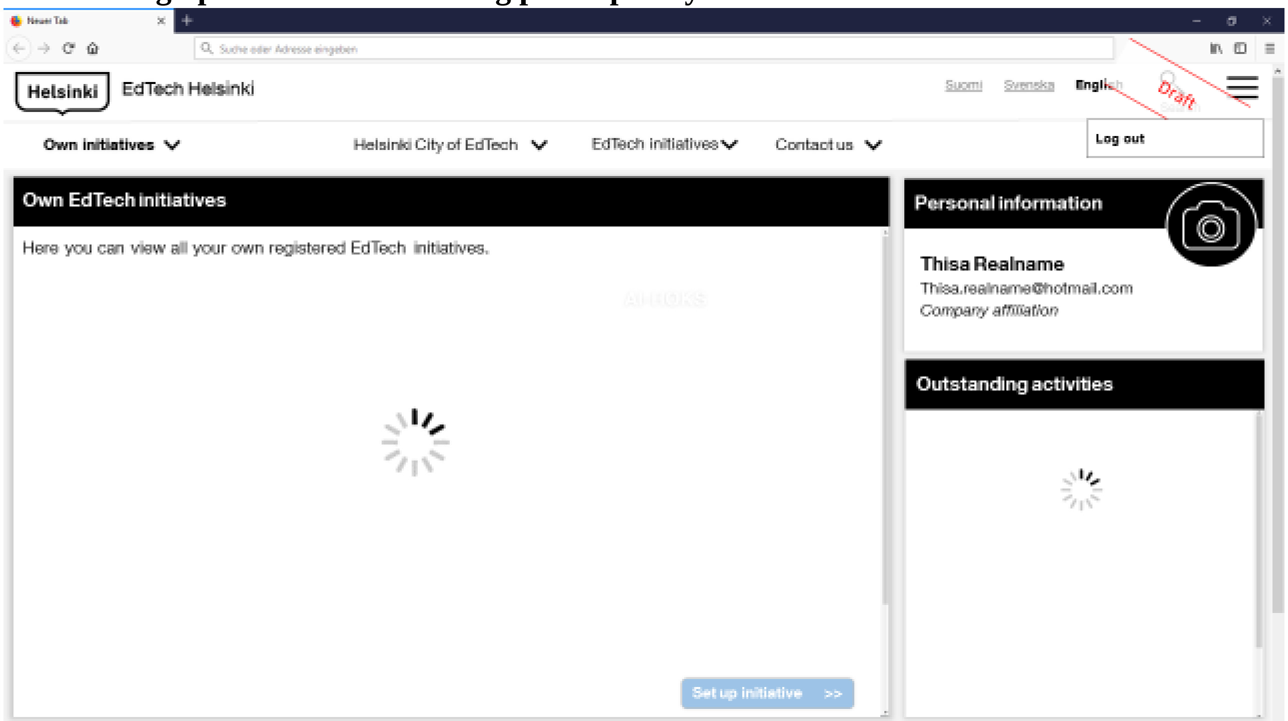


Figure 14: Overview page of EdTech launcher

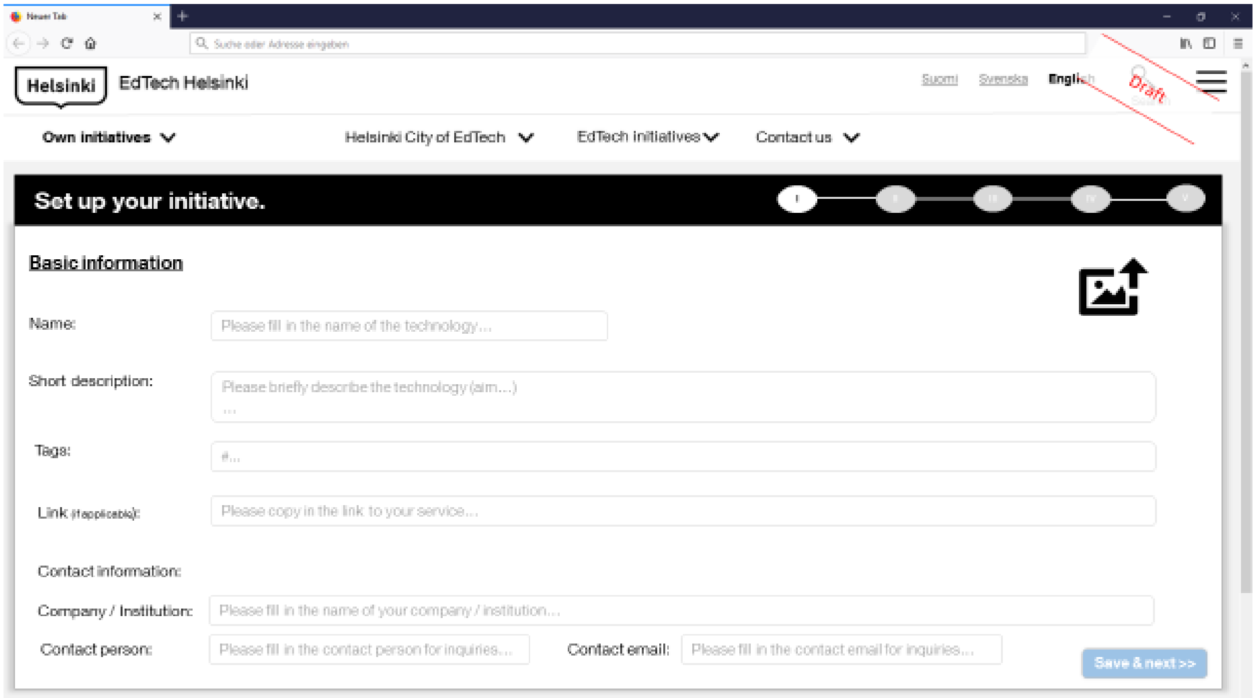


Figure 15: First step – Set up & fill in basic information

First, EdTech launcher users need to fill in basic information* about their initiative (Figure 15).

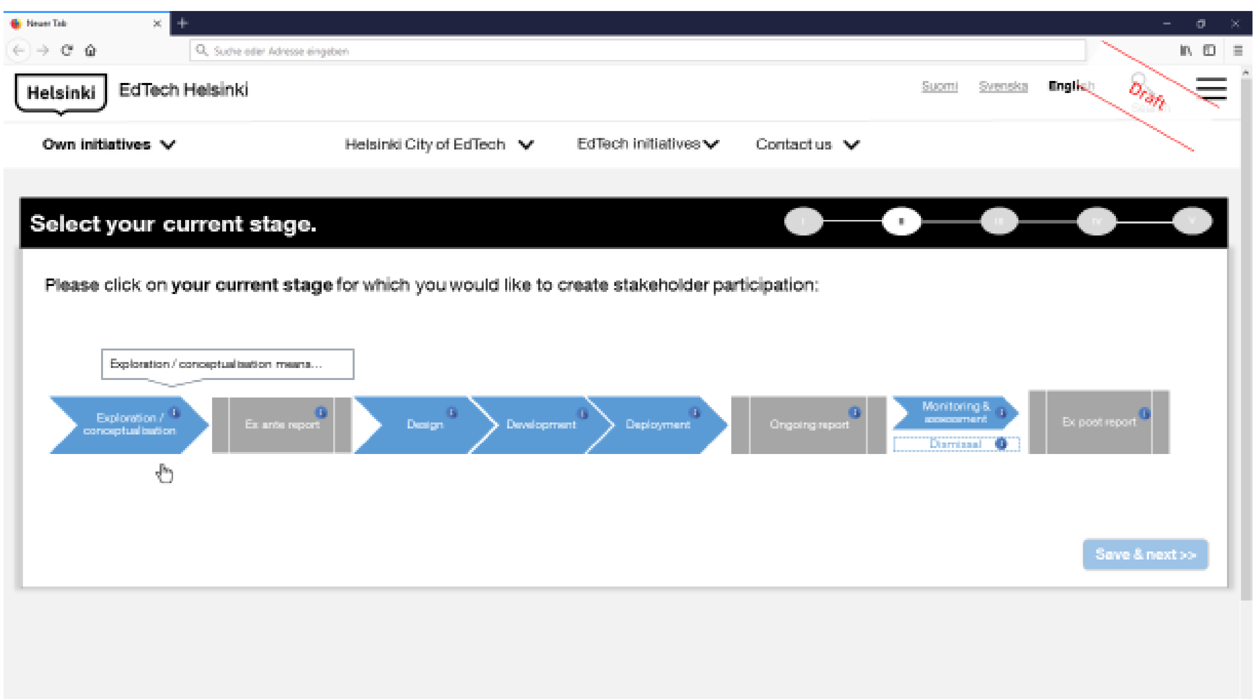


Figure 16: Second step – Select current stage

Second, EdTech launcher users select their current stage within the technology life cycle** (Figure 16).

Third, EdTech launcher users can start composing their own participatory toolkit (Figure 17 & Figure 18).

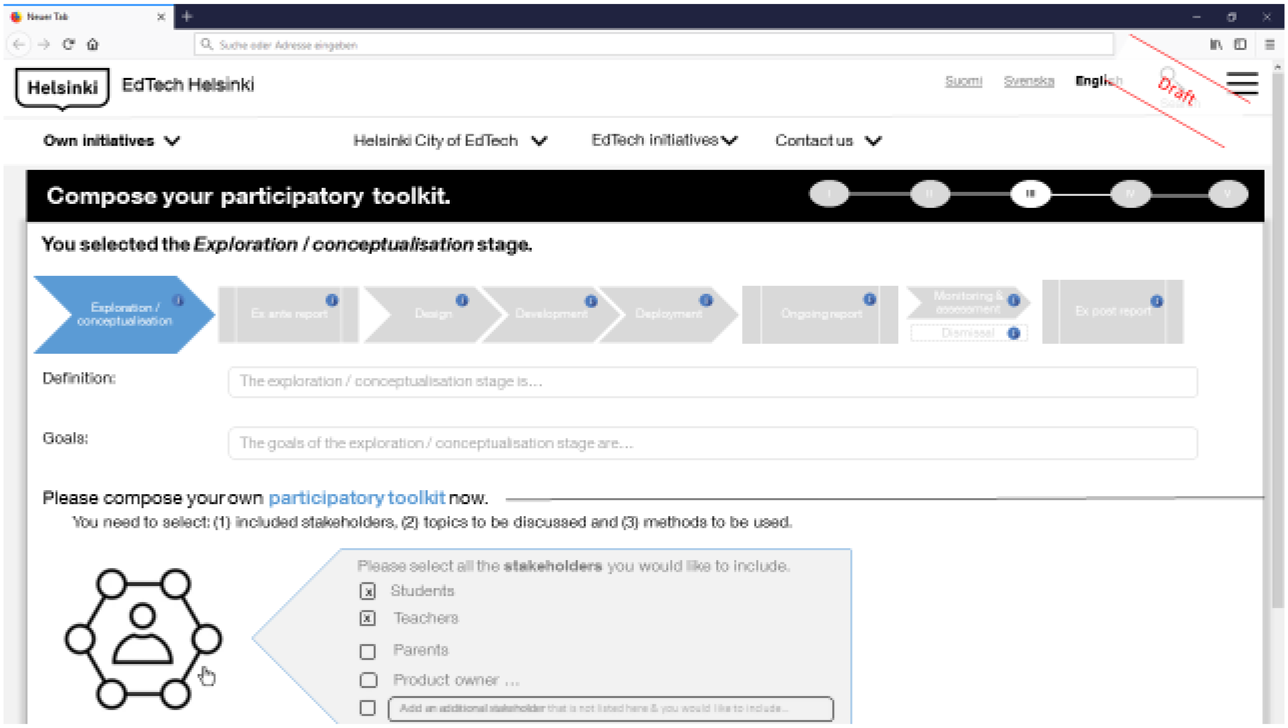


Figure 17: Third step – Compose participatory toolkit (1)

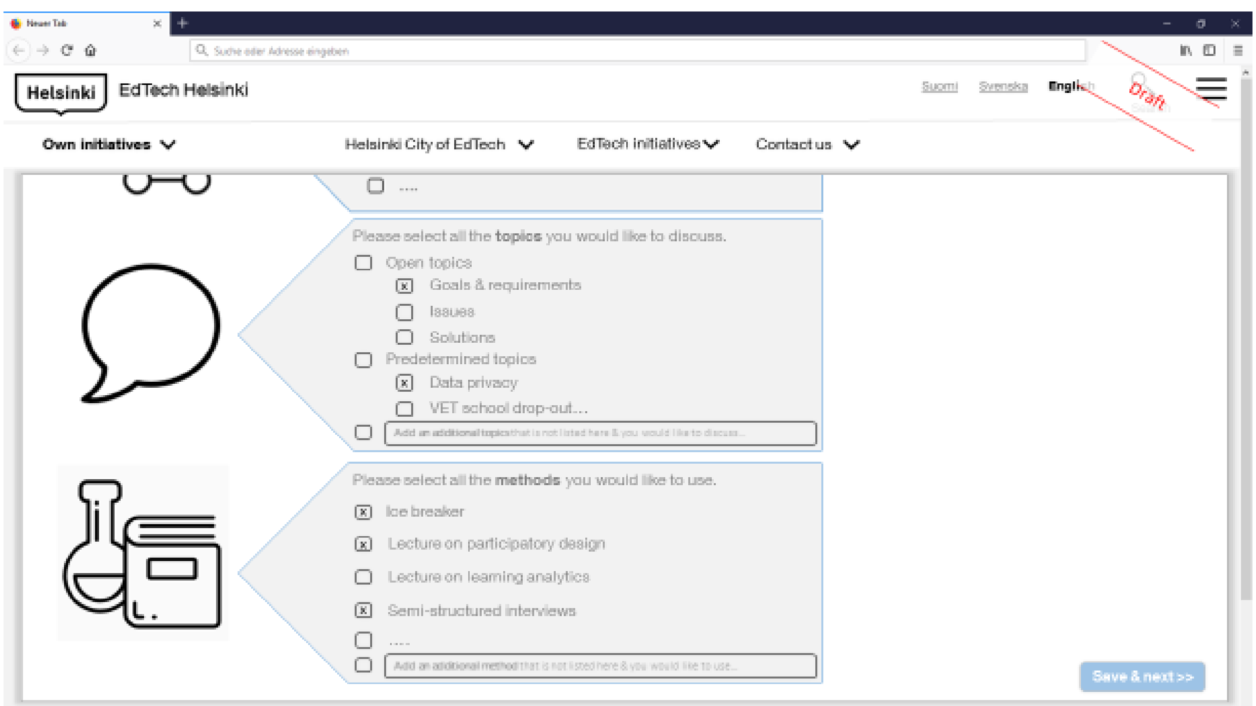


Figure 18: Third step – Compose participatory toolkit (2)

In particular, users here need to choose stakeholders that should be included, topics that should be discussed as well as methods to be utilized. The idea would be that the options that are displayed here are relevant for the particular stage within the technology life cycle***. For example, it is necessary to include some particular stakeholder, topics and methods in the exploration stage while it is necessary to include other particular stakeholder, topics and methods in the final development stage. In addition to predetermined options, EdTech launcher users themselves can add stakeholders, topics and methods****.

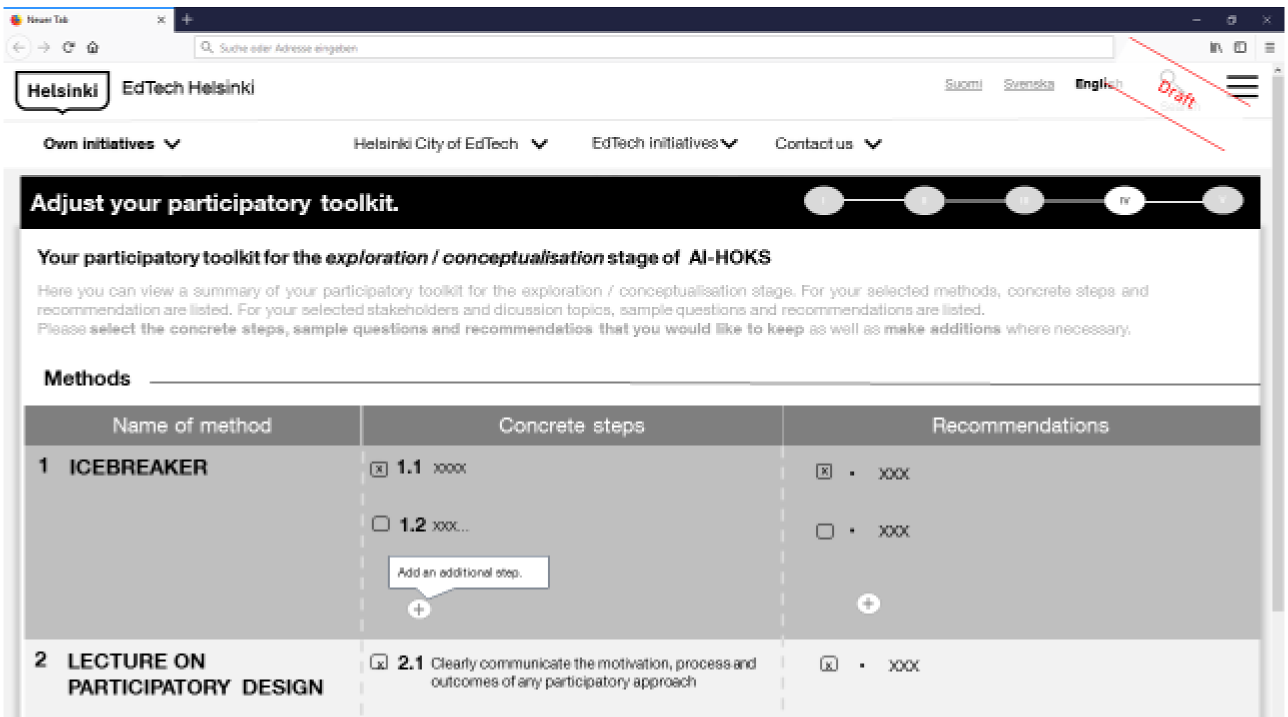


Figure 19: Fourth step – Adjust participatory toolkit (1)

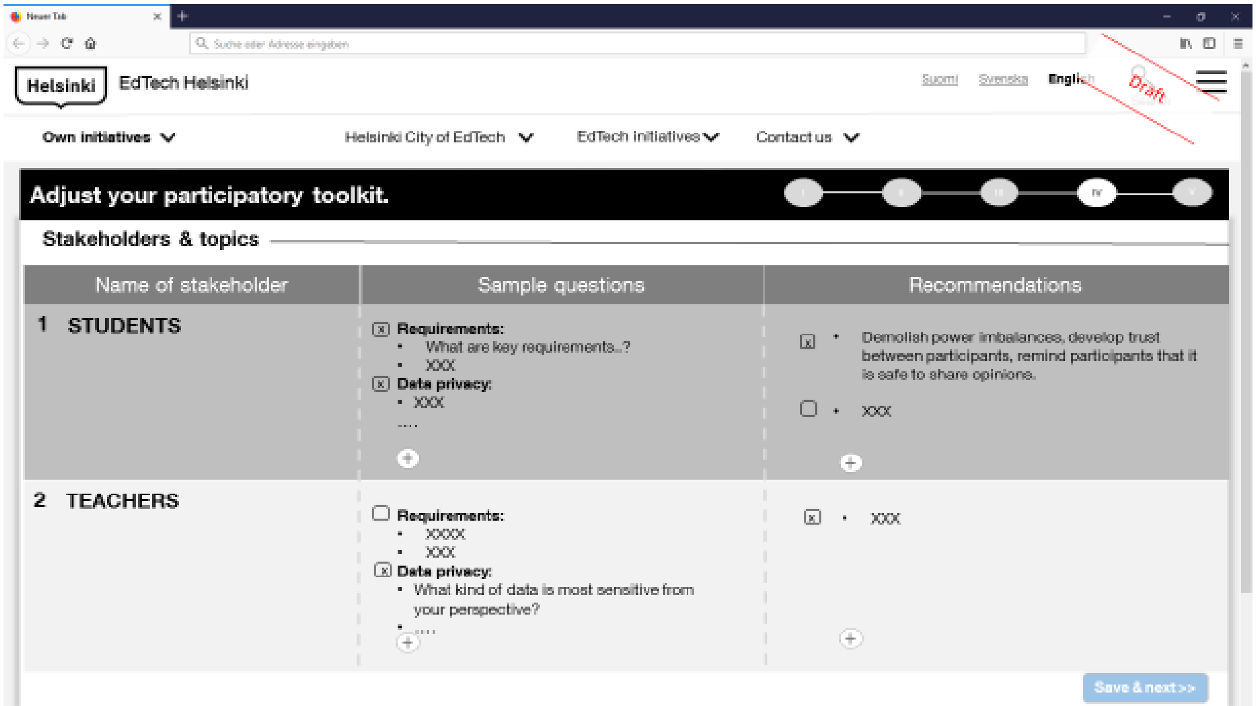


Figure 20: Fourth step – Adjust participatory toolkit (2)

Fourth, EdTech launcher users can view and adjust their previous selection for their own participatory toolkit. What will be displayed is the methods that they have selected in step three. For each method, here, concrete steps and recommendations will be displayed***** (Figure 19). Furthermore, what will be displayed is the stakeholders they selected in step three. For each stakeholder, here, sample questions and recommendations will be displayed******(Figure 20). In addition to the predetermined concrete steps, sample questions, and recommendations, EdTech launcher users themselves can make their own additions****.

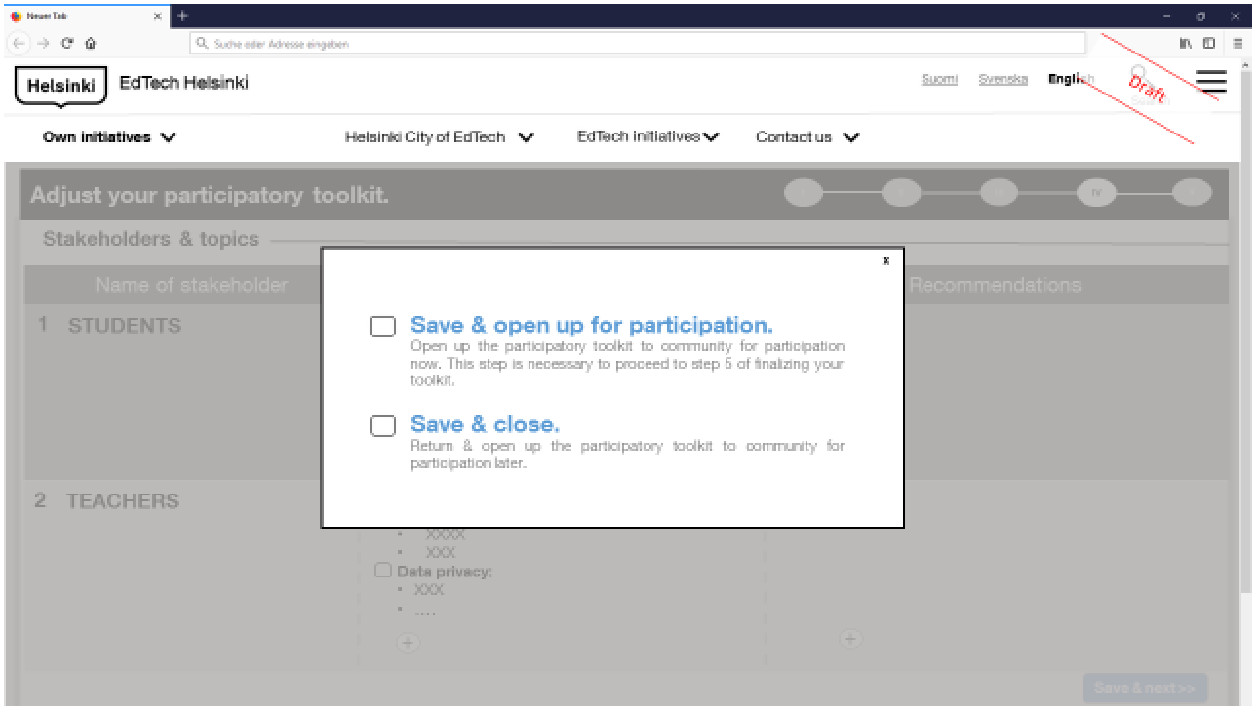


Figure 21: Fourth step – Save & open up for participation

Following this, EdTech launcher users can save their adapted participatory toolkit and open it up for participation. This means, once "Save & open up for participation" is selected, the initiative will be visibly listed in the initiative register (Figure 31) and the community can start viewing and making additions to this initiative, i.e. to the created participatory toolkit (e.g. Figure 33 & Figure 35).

Furthermore, the initiative will be listed on the overview page of the corresponding EdTech launcher user (Figure 22). Here the EdTech launcher can also track whether the community has conducted evaluations or signed up to participate. After having the participatory toolkit open for participation for some time and when the EdTech launcher user decides that enough community evaluations have been generated, the EdTech launcher can close the ability to participate******* and finalize the participatory toolkit by clicking on the corresponding button.

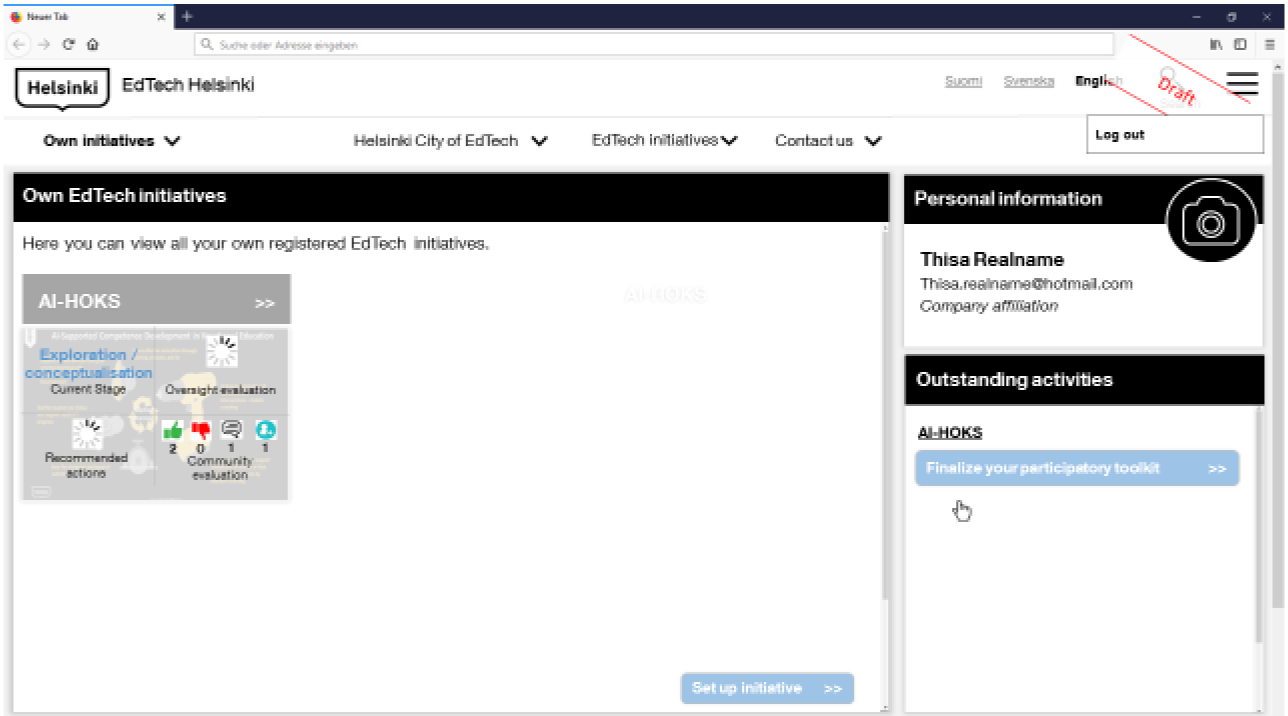


Figure 22: Updated overview page of EdTech launcher – before finalization of participatory toolkit

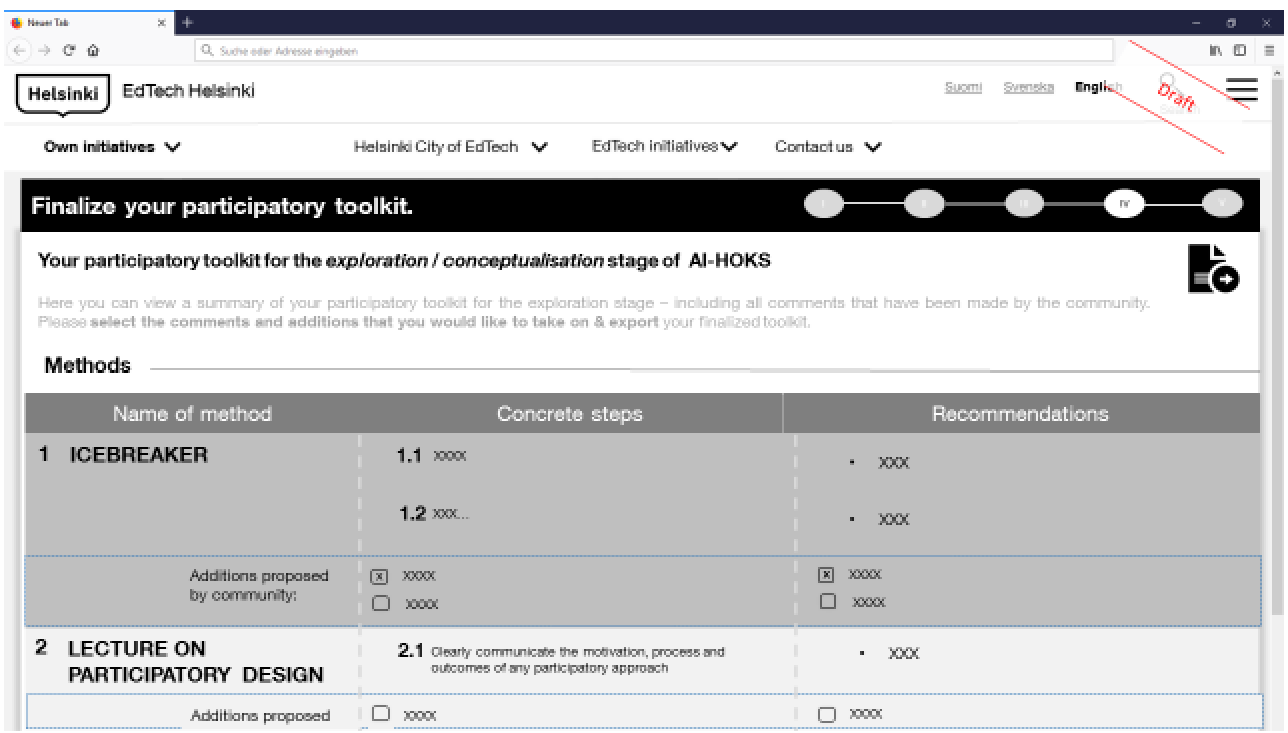


Figure 23: Fifth step – Finalize participatory toolkit (1)

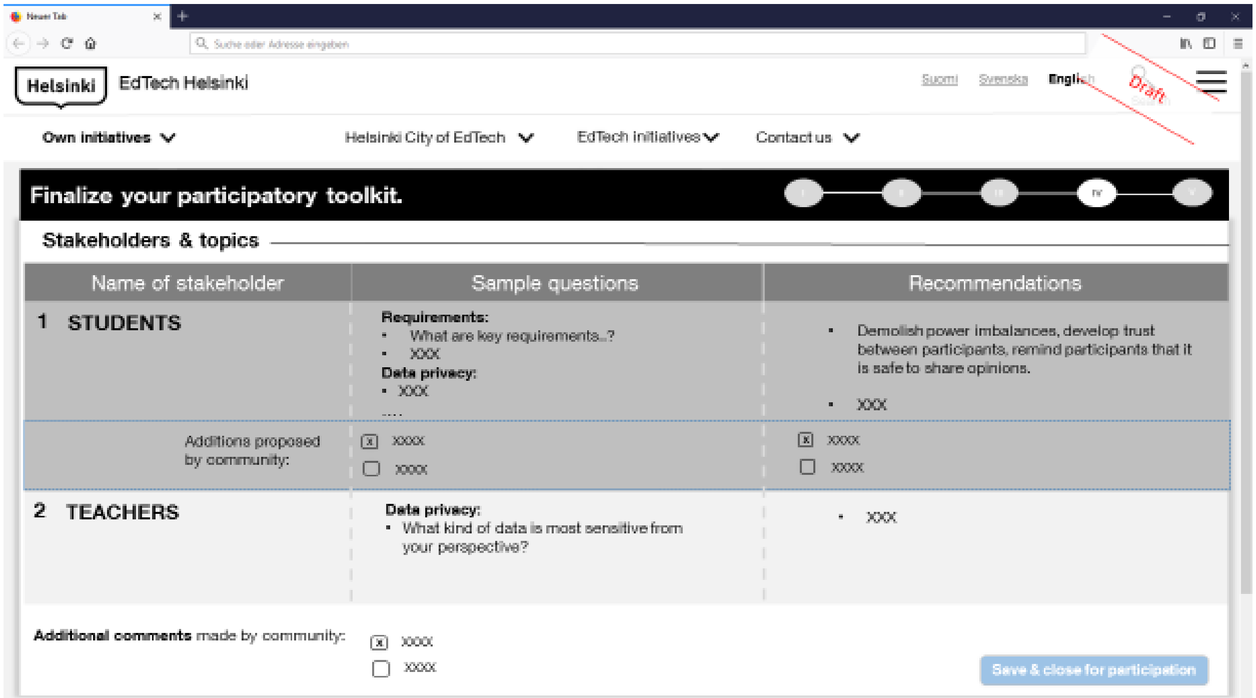


Figure 24: Fifth step – Finalize participatory toolkit (2)

Fifth, the EdTech launcher user can view all the additions that were proposed by the community. The EdTech launcher then can select which of the additions by the community he/she would like to keep or disregard. As a result, the EdTech launcher user has generated a finalized participatory toolkit which he/she can download, go ahead and use it for conducting his/her workshop, interview etc. (Figure 23 & Figure 24).

Limitations / things to consider:	• *The layout of the "basic information" form should be similar to the City of Helsinki AI Register to allow smooth integration (more information in 3.2 Suggestions for integration) • ****In case EdTech launcher users have added additional stakeholders, topics and methods, this information should be transferred to the host of the platform. It should be then verified whether the addition is an important entry that has to be added to the overall back-end so that this particular entry will be displayed to all users in the future (more information in 3.1 Recap of key features). • *****While the Policy Paper provides municipal stakeholder engagement strategies for learning analytics and AI in Education on a general level,

	concrete steps and recommendations that can be stated in this wireframe still need to be developed. ● *******As stated, the EdTech launcher can decide how long the participatory toolkit is open for participation for the community. To ensure that the community has sufficient time to co-design, the City of Helsinki should set a minimum amount of time in which the EdTech launcher cannot close this stage.
Relevant back-end material:	● **For more information, see “Model for Stakeholder Engagement”. ● ***For more information, see “Translational Models”: While the stated matrix does not link particular participatory methods to particular stages of the technology life cycle, it can help identify important topics and stakeholders to be included and according methods to be utilized. ● ******For more information (i.e., sample questions) see “Key Stakeholders in Education”.

2.2.2 Reporting

As stated in Figure 16, EdTech launchers need to submit reports at three different stages during the technology life cycle: ex ante report, ongoing report and ex post report.

For example, once the EdTech launcher of the AI-HOKS initiative generated a finalized participatory toolkit and used it for conducting his/her workshop, interview etc., the EdTech launcher needs to submit an ex ante report by clicking on the corresponding button (Figure 25).

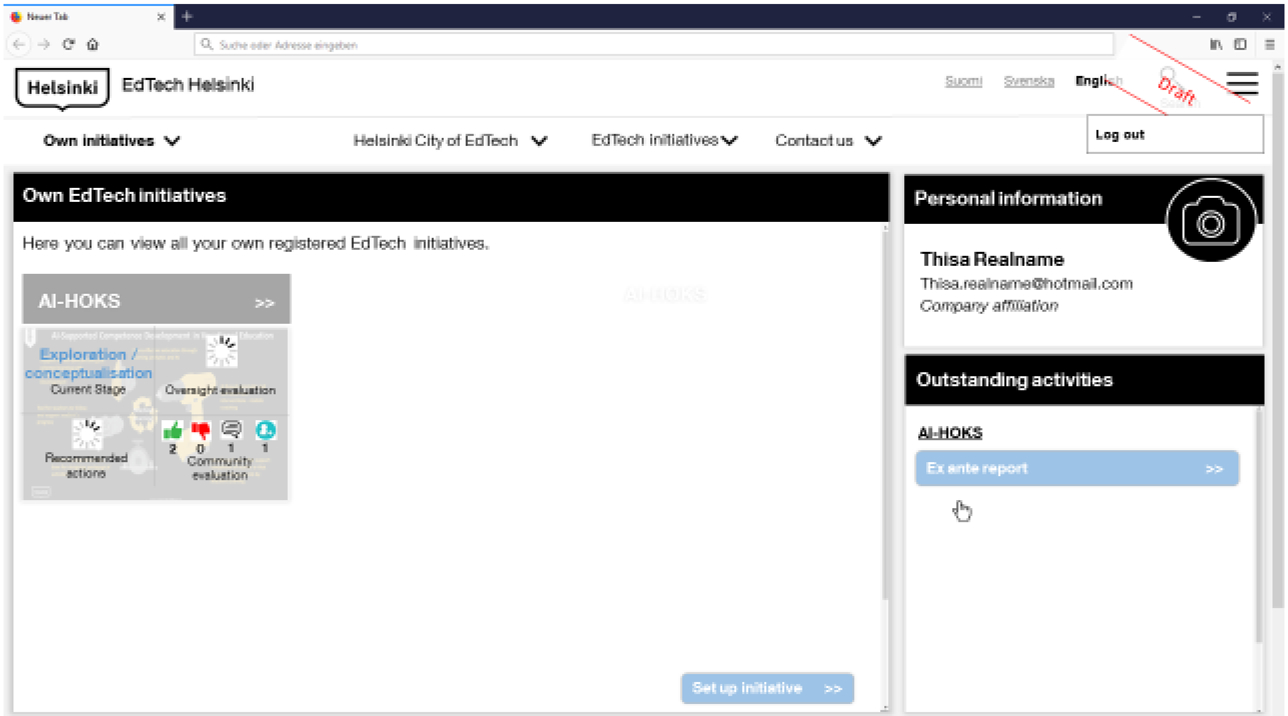


Figure 25: Updated overview page of EdTech launcher – before ex ante report

When conducting the ex ante report, the EdTech launcher has to elaborate on the insights generated during the previous stage. Namely, they need to state their lessons learned and consequent adaptations (Figure 27) as well as conduct an ethics & risk self-assessment* by reporting their actions taken to fulfill certain ethical requirements (Figure 27 & Figure 28).

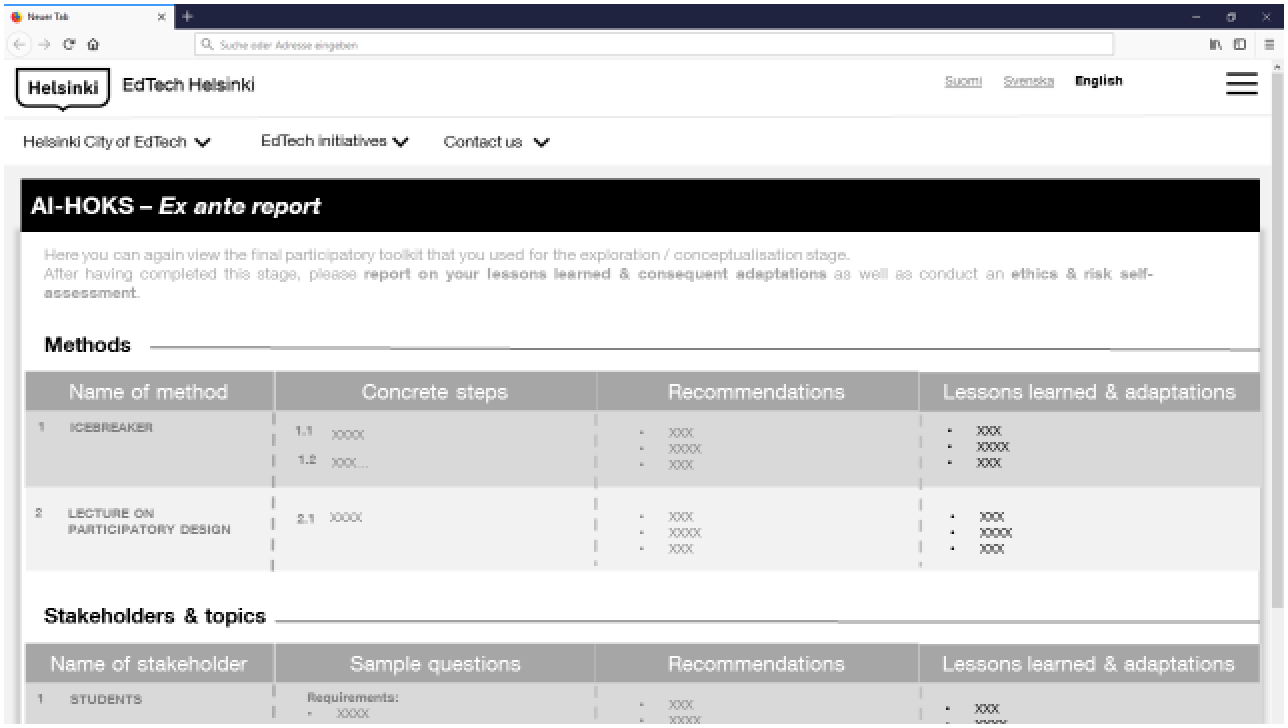


Figure 26: Ex ante report (1)

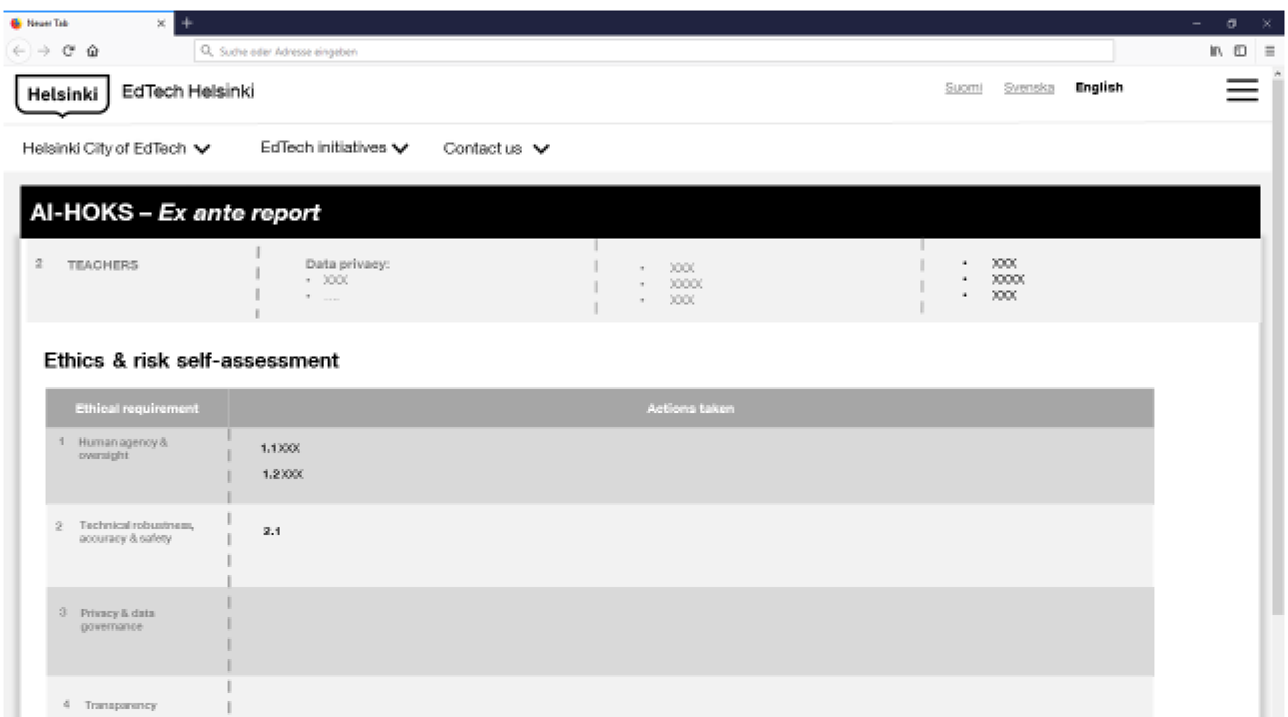


Figure 27: Ex ante report (2)

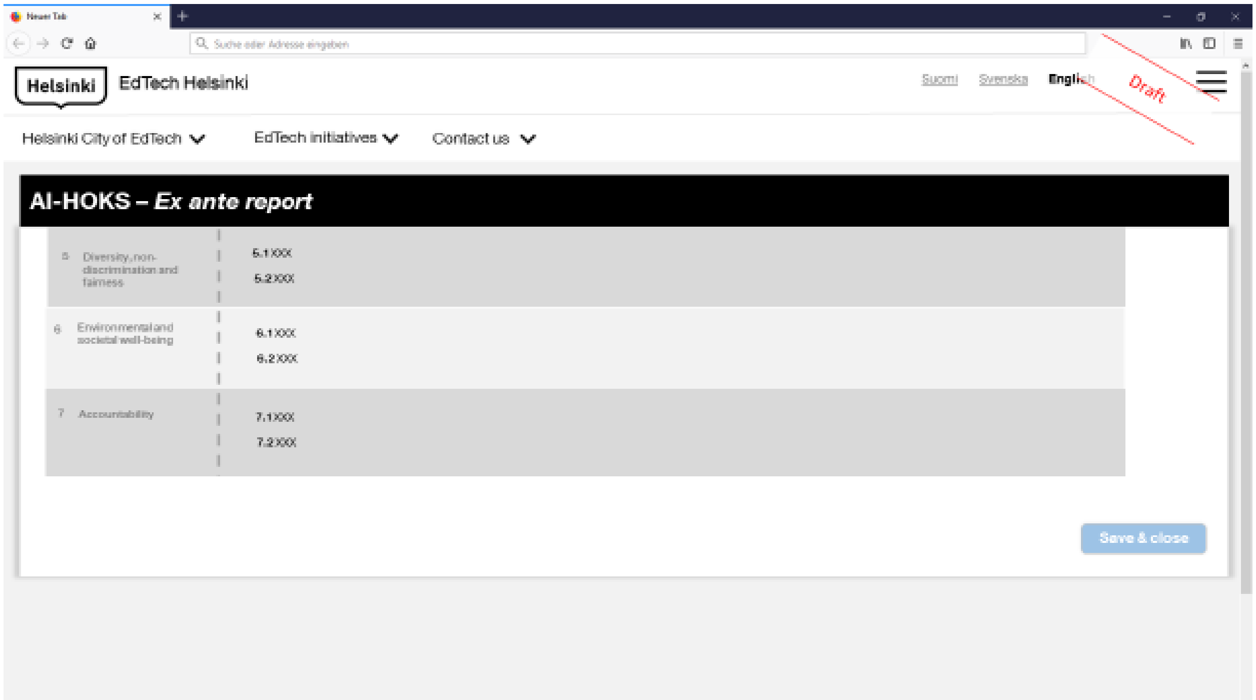


Figure 28: Ex ante report (3)

Once the EdTech launcher has saved and closed the ex ante report, the report is available for feedback to the community (Figure 37 & Figure 38) as well as for evaluation to the oversight team members (Figure 44 & Figure 45).

The ex ante report and the ongoing report are equal in terms of the content that needs to be reported on. However, the ex post report that needs to be conducted after the monitoring & assessment stage differs in terms of content as illustrated in the following.

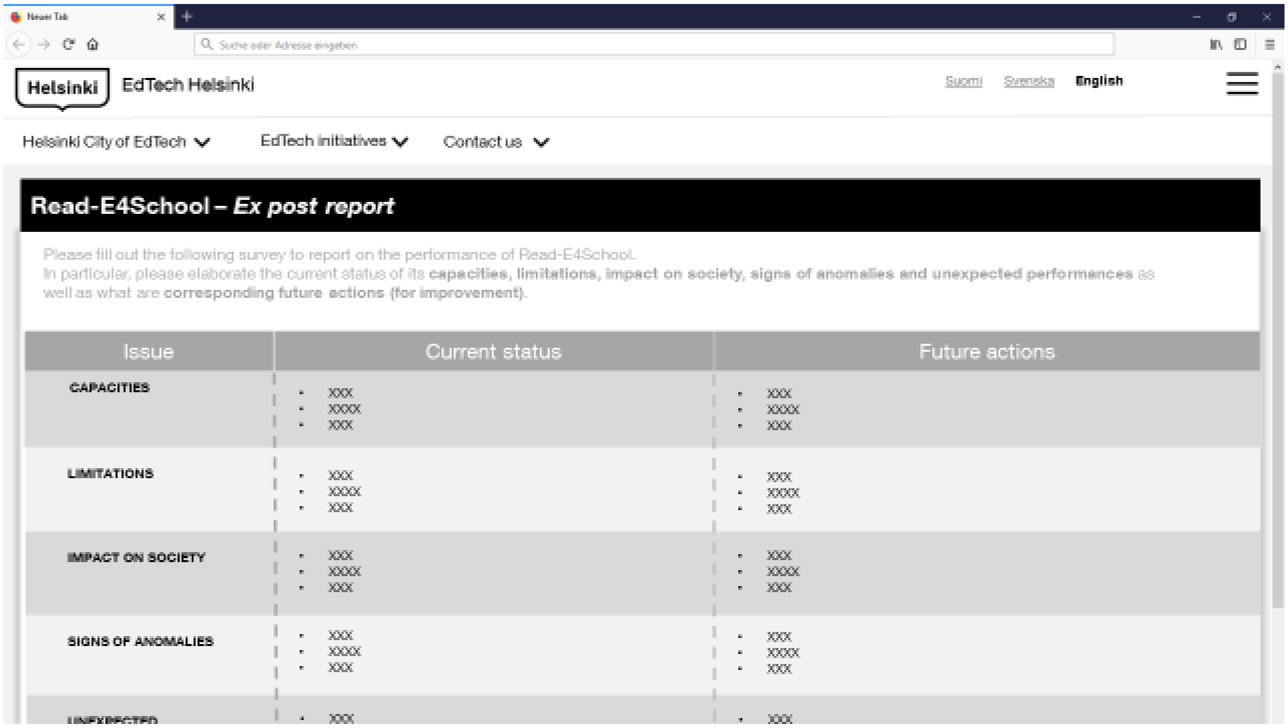


Figure 29: Ex post report (1)

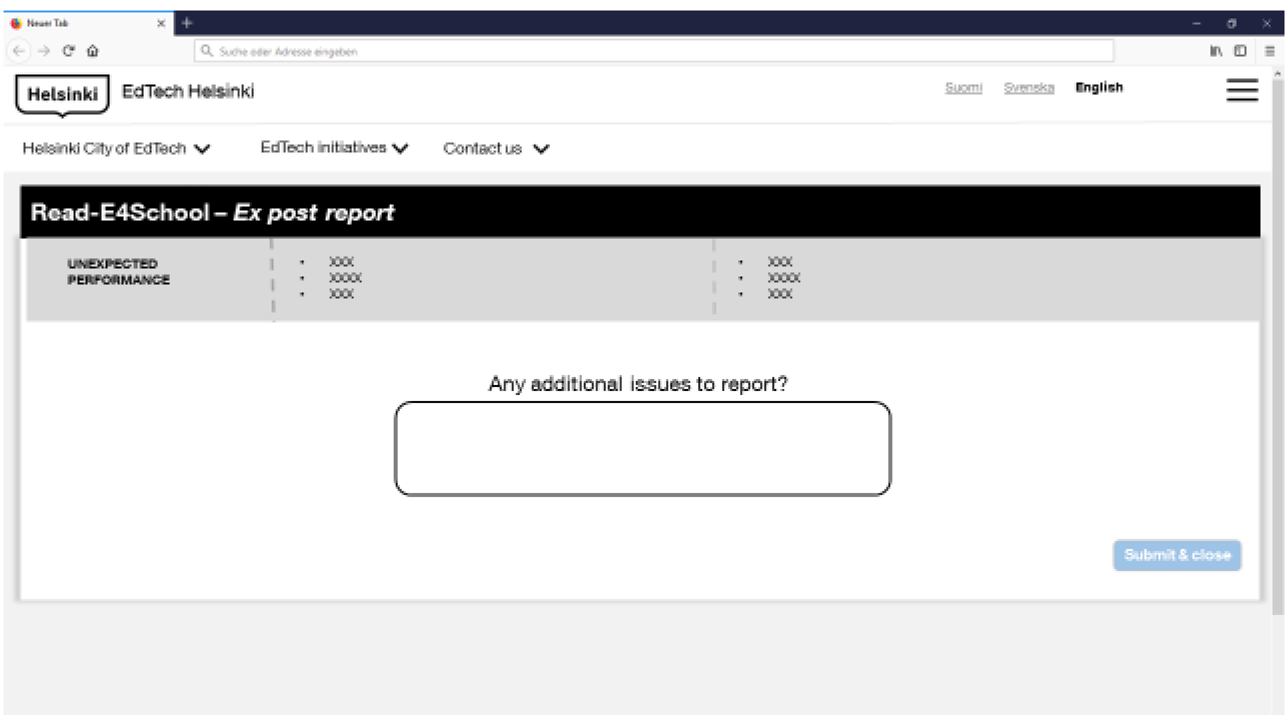


Figure 30: Ex post report (2)

Within the ex post report, EdTech launchers need to elaborate on the current status of the technology's capacities, limitations, impact on society, signs of anomalies and unexpected consequences as well as what are corresponding future actions (for improvement)** (Figure 29 & Figure 30). Once the EdTech launcher has saved and closed the ex post report, the report is available to the community as well as for evaluation to the oversight team members (Figure 47).

Relevant back-end material:	● *For more information, see "Model for Stakeholder Engagement" and "Translational Models"; The theoretical ground for ethics & risk self-assessment can be found in the Matrix, which was heavily based on the Assessment List for Trustworthy Artificial Intelligence of the High-level Expert Group on AI (2019). ● ** For more information, see Appendix I: AI Ethics systemic translational matrix.

2.3 Website flow for community users

On the tab “EdTech initiatives”, users can view all EdTech initiatives that have been created / listed on this platform (Figure 31).

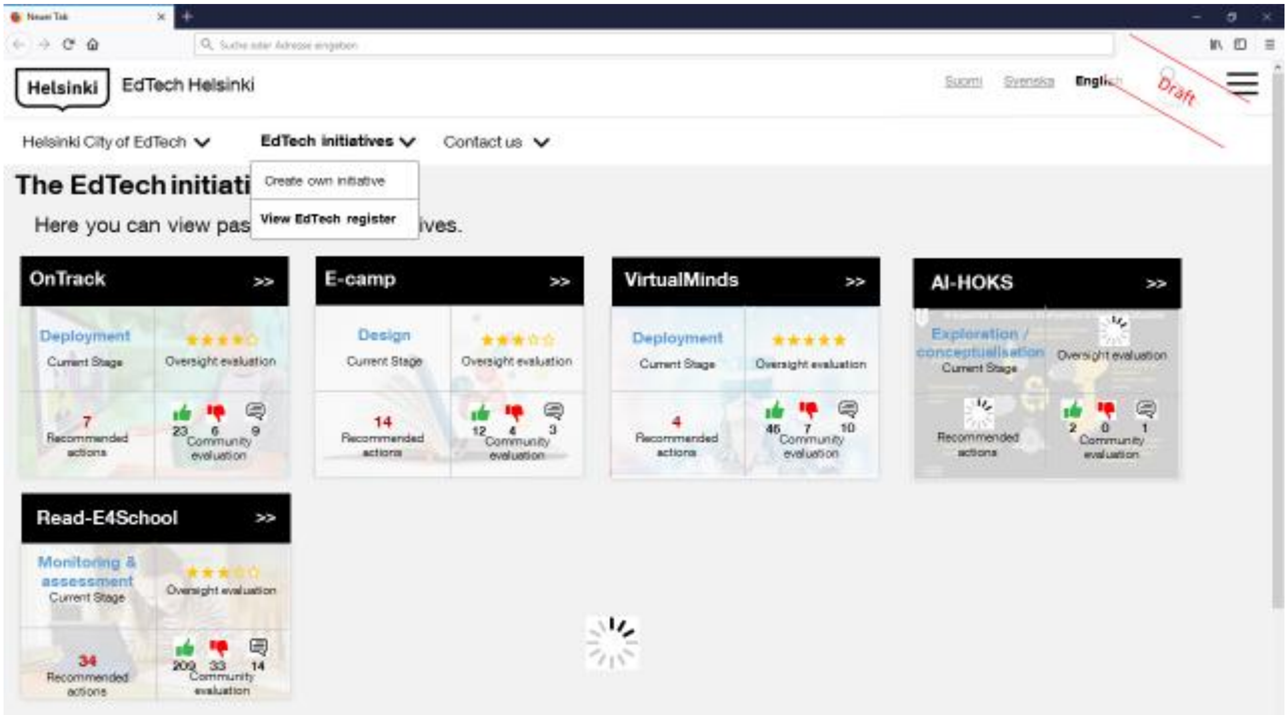


Figure 31: Viewing the EdTech initiatives Register

By clicking on a particular initiative, community users can view more detailed information (Figure 32).

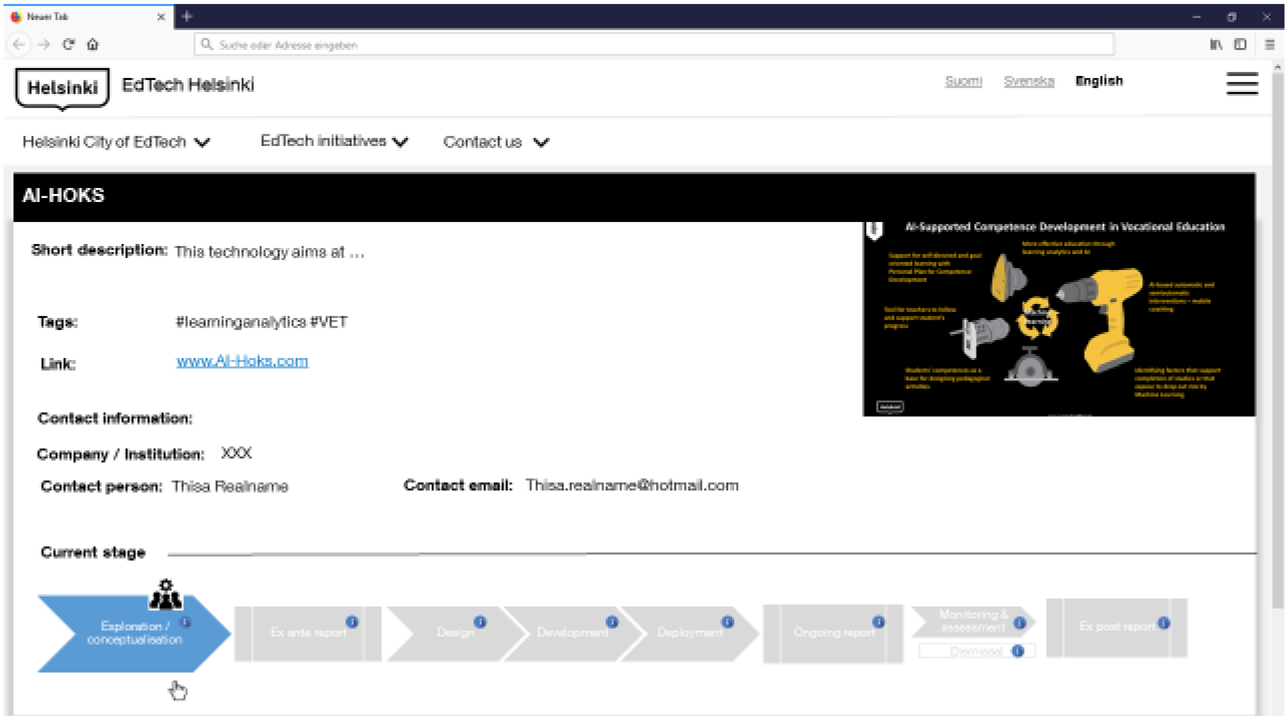


Figure 32: Detailed information about particular EdTech initiative AI-HOKS (for community users) (1)

The detailed information that will be displayed here is the information that the responsible EdTech launcher had filled in when setting up the initiative (Figure 15). In addition, the community user can view the current stage of the particular initiative and provide corresponding feedback*. For example, here the particular initiative is located at the "Exploration / conceptualization" stage (Figure 32). For this current stage, community users can co-design/ provide ad hoc comments by clicking on  (Figure 33, Figure 34 & Figure 35).

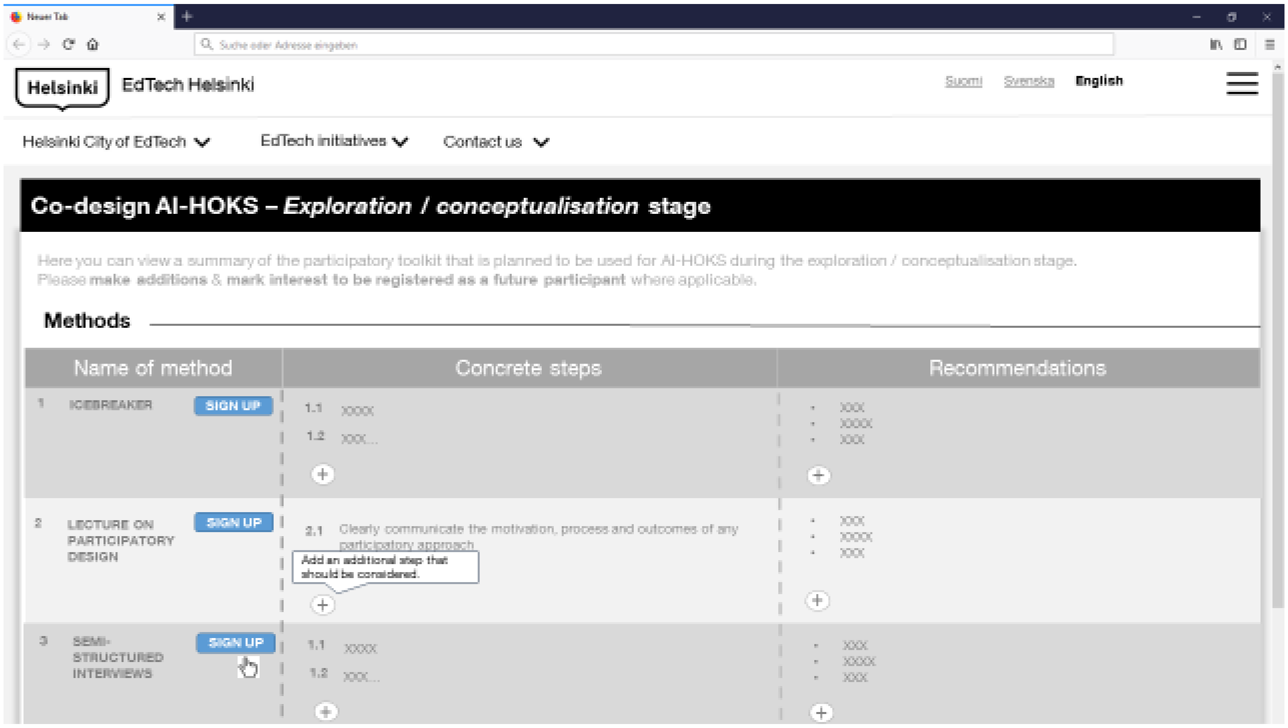


Figure 33: Co-design & ad hoc comments (1)

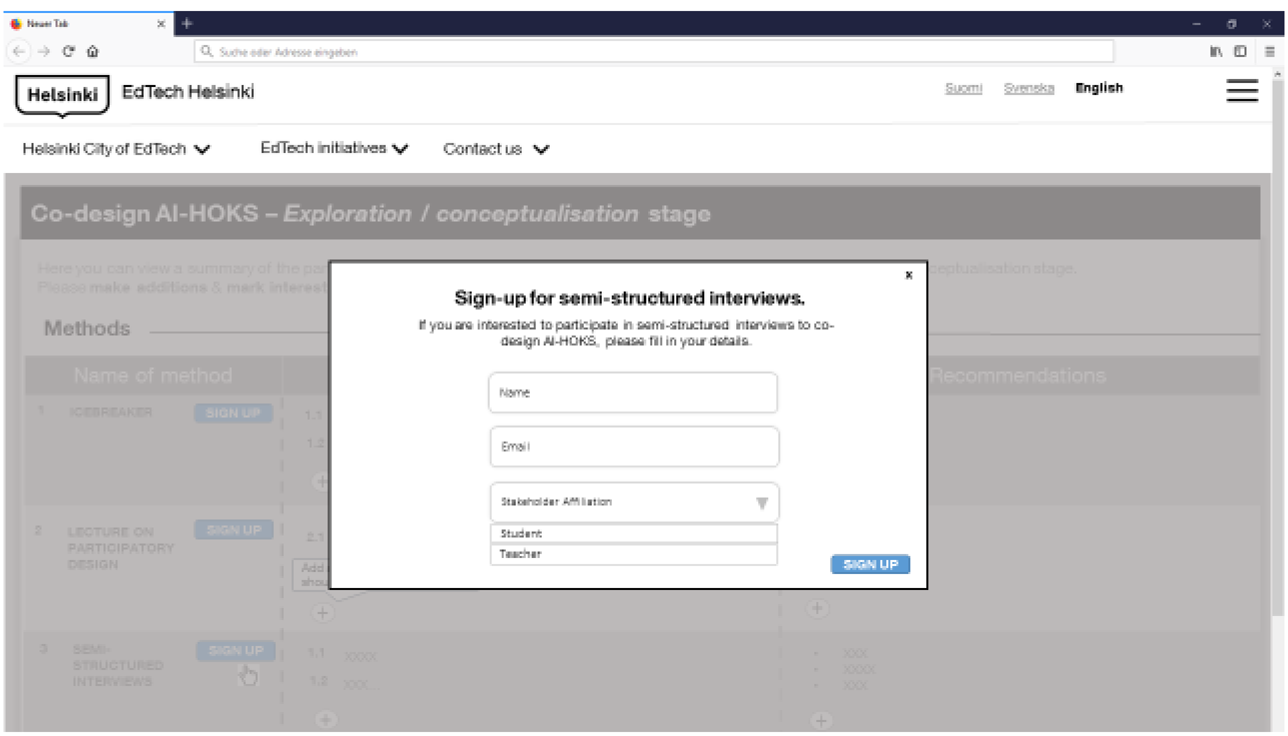


Figure 34: Co-design & ad-hoc comments (2)

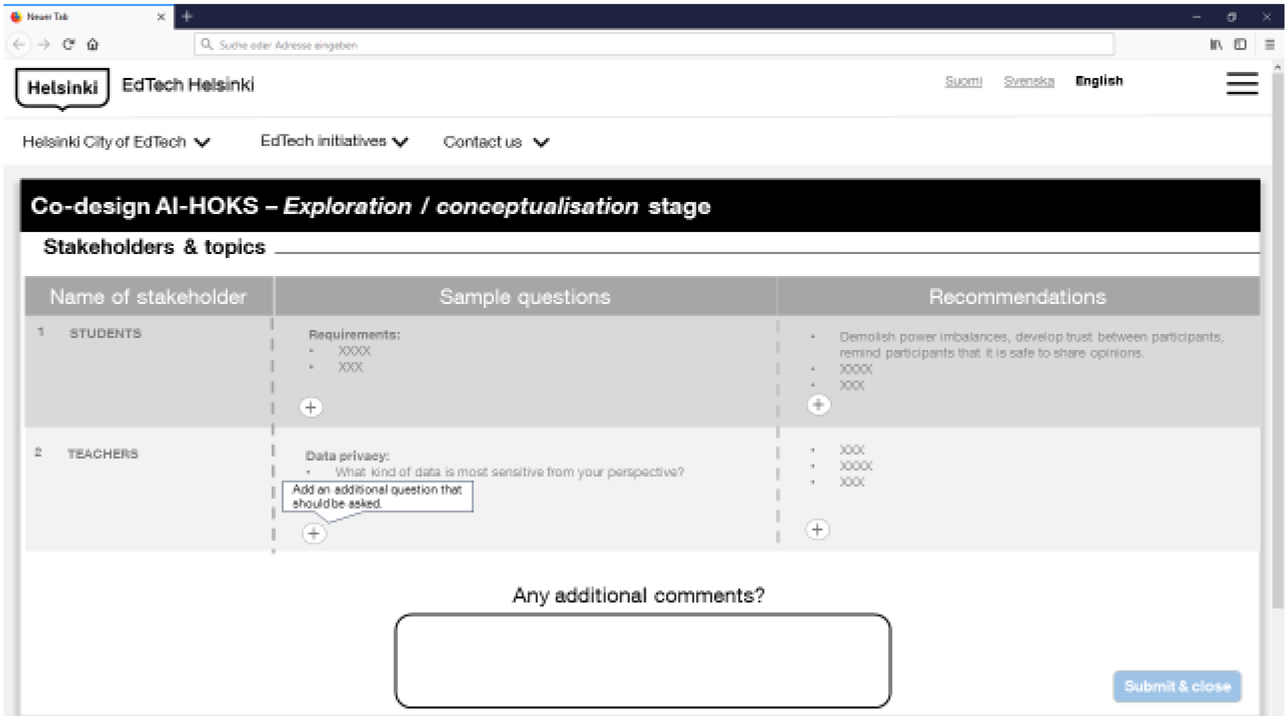


Figure 35: Co-design & ad hoc comments (3)

When giving ad hoc feedback (e.g. in this example to the final development stage), community users can view the adjusted participatory toolkit that was published by the EdTech launcher (Figure 21), propose to add concrete steps, sample questions and recommendations, provide additional comments as well as sign up for a particular method if interested (Figure 33, Figure 34 & Figure 35). All this information filled in by community users will be transferred back to the EdTech launcher (Figure 23 & Figure 24). Furthermore, the number of comments etc. by the community users will be taken into account and aggregated in the "community evaluation" (Figure 42).

In addition to co-design/ ad hoc feedback, community users have the ability to provide ex post feedback to past stages. For example, if the particular initiative is located at the "Design" stage, all previous stages are finished as indicated by the green checkmark (Figure 36). For all the finished stages, community users can provide ex post feedback to a certain stage by clicking on the corresponding  (Figure 37, Figure 38, Figure 39 & Figure 40).

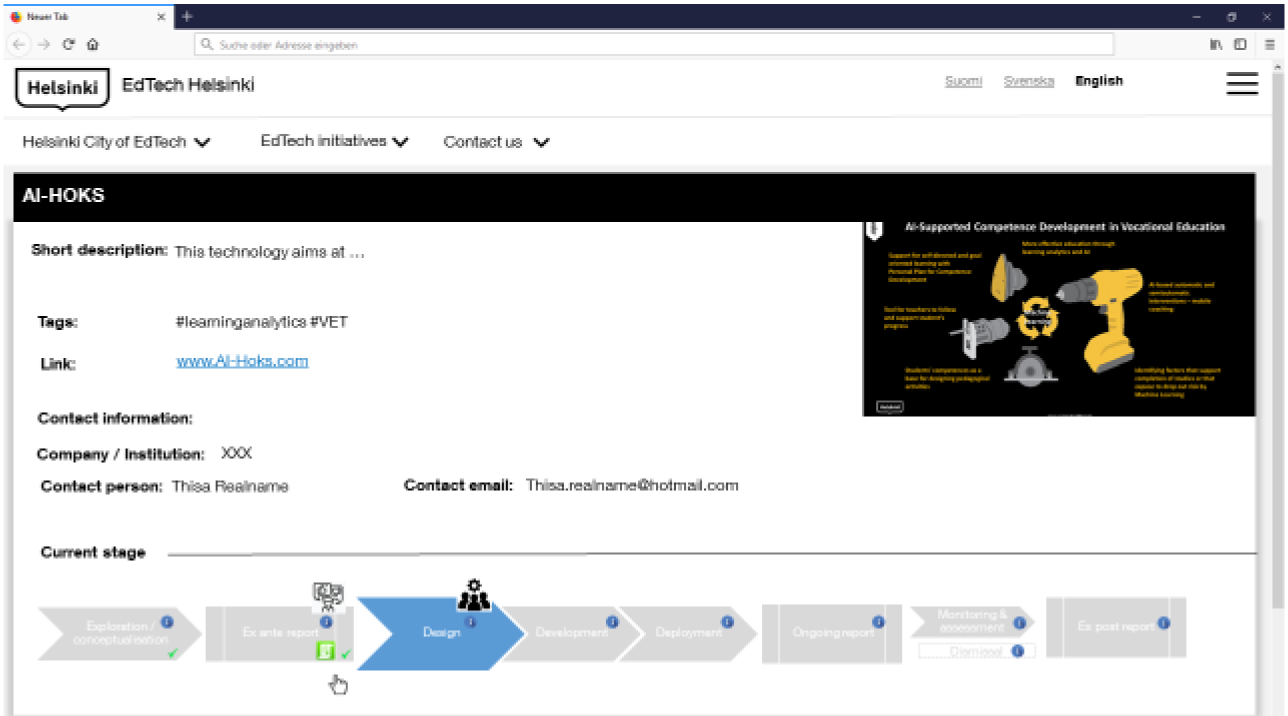


Figure 36: Detailed information about particular EdTech initiative AI-HOKS (for community users) (2)

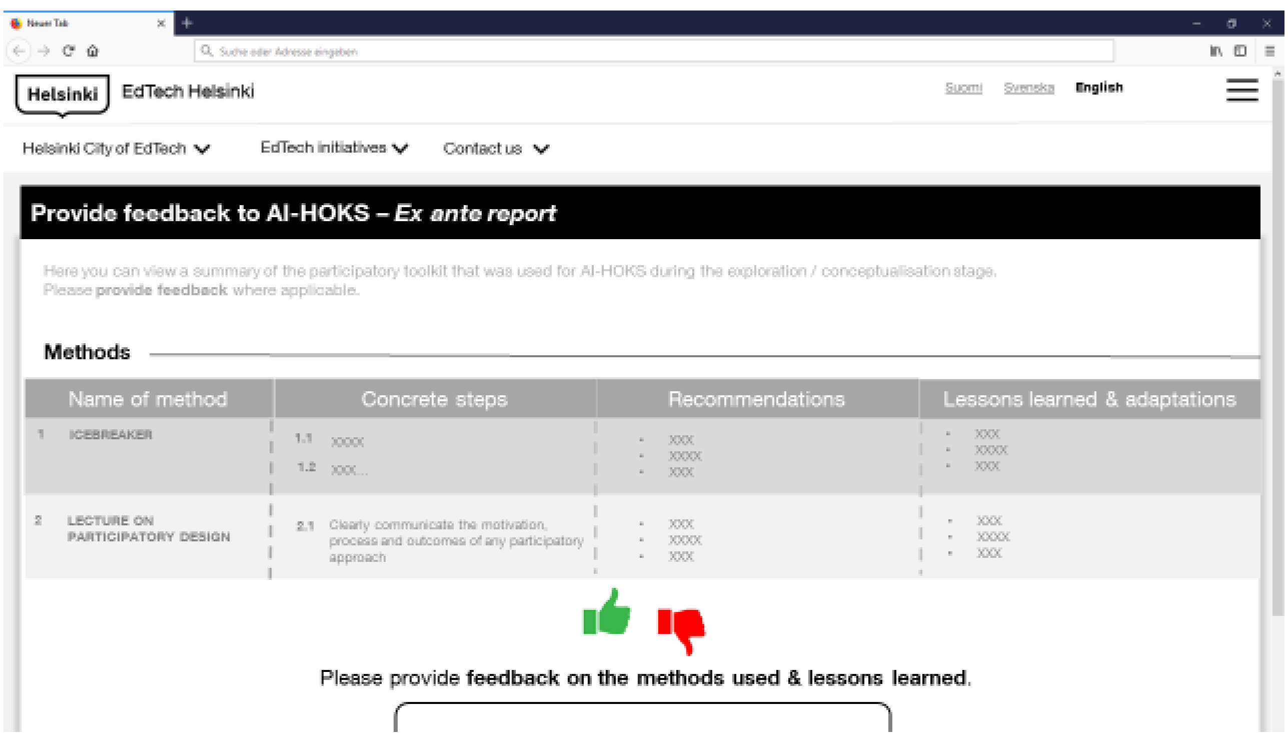


Figure 37: Ex post feedback (1)

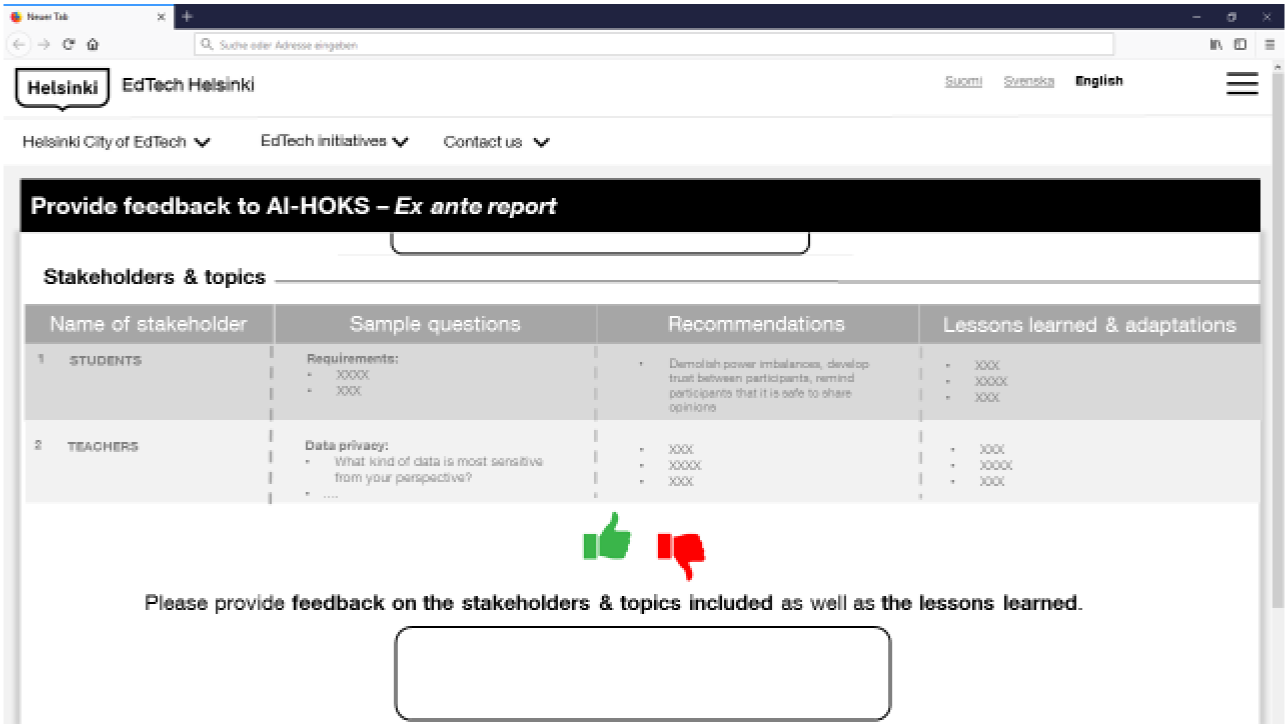


Figure 38: Ex post feedback (2)

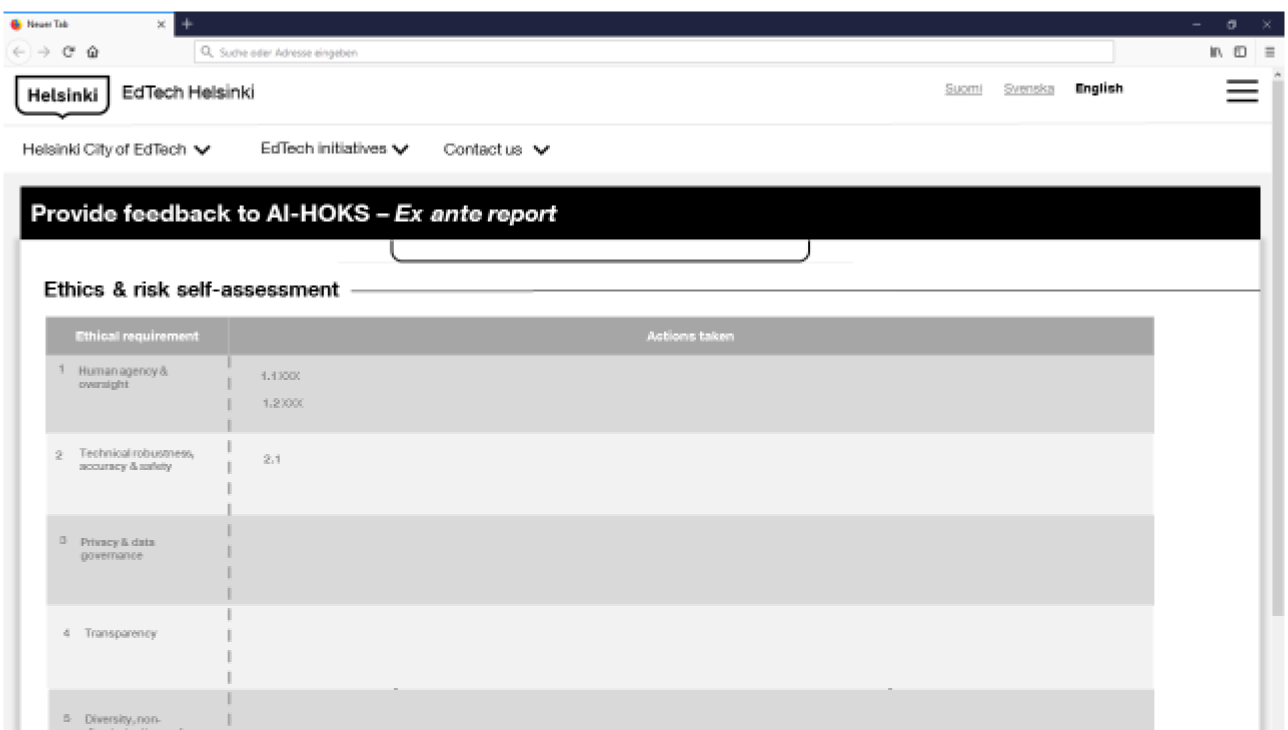


Figure 39: Ex post feedback (3)

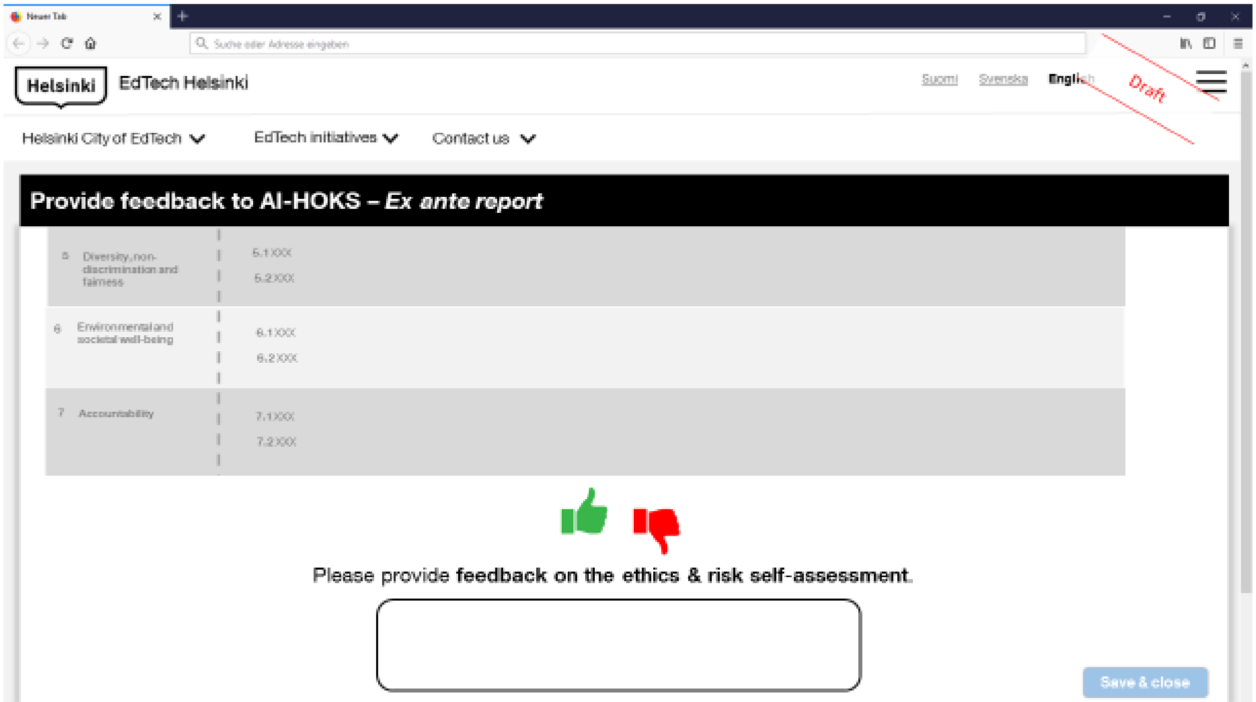


Figure 40: Ex post feedback (4)

When giving ex post feedback, community users can view the finalized participatory toolkit that was used by the EdTech launcher as well as any lessons learned and consequent adaptations that the EdTech launcher developed after conducting his/her workshop, interview etc. (Figure 37 & Figure 38). In addition, community users can view the ethics & risk self-assessment that the EdTech launcher created (Figure 39 & Figure 40). For the methods utilized, the stakeholders and topics included as well as the ethics & risk self-assessment, community users can like or dislike the stated actions. In addition, community users can provide written comments. The number of likes, dislikes and comments by the community users will be taken into account and aggregated in the "community evaluation" (e.g., Figure 22, Figure 31 & Figure 42).

Relevant back-end material:	• *For more information, see "Model for Stakeholder Engagement".

2.4 Interface for oversight team

Oversight team members can use the platform by logging onto the oversight team interface (Figure 41).

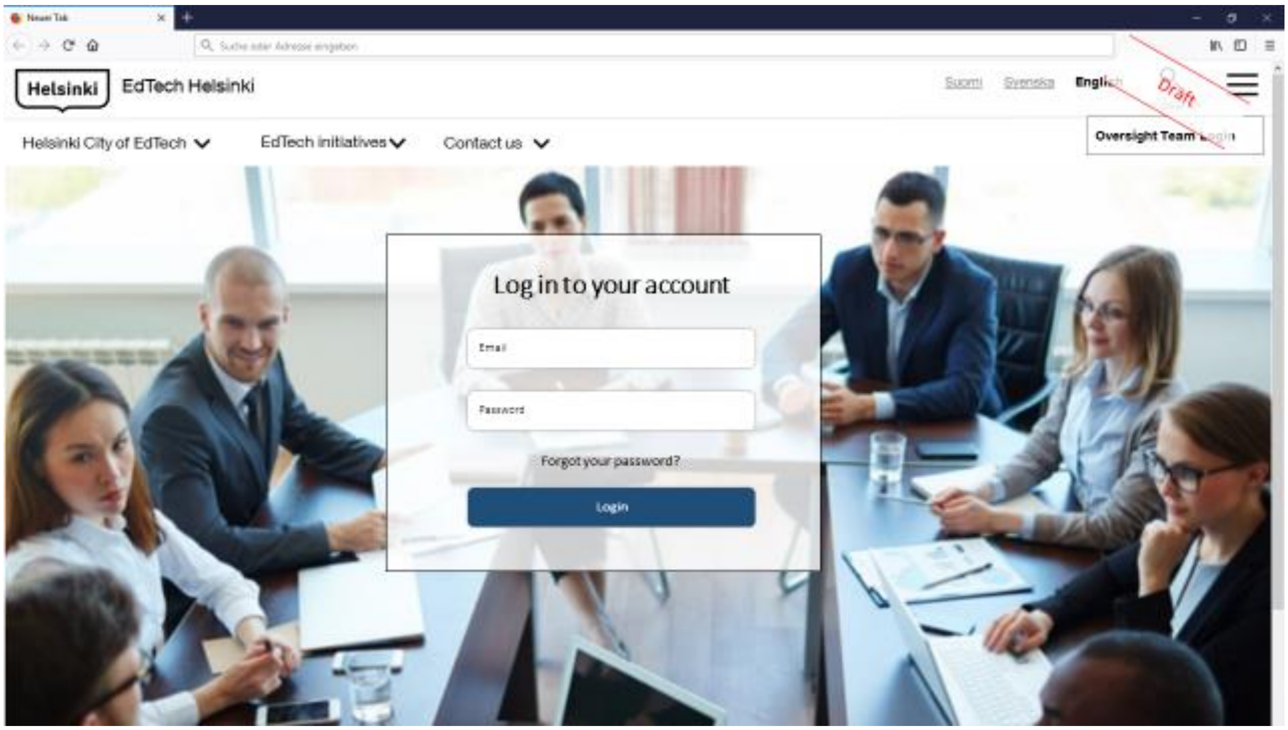


Figure 41: Login of oversight team members

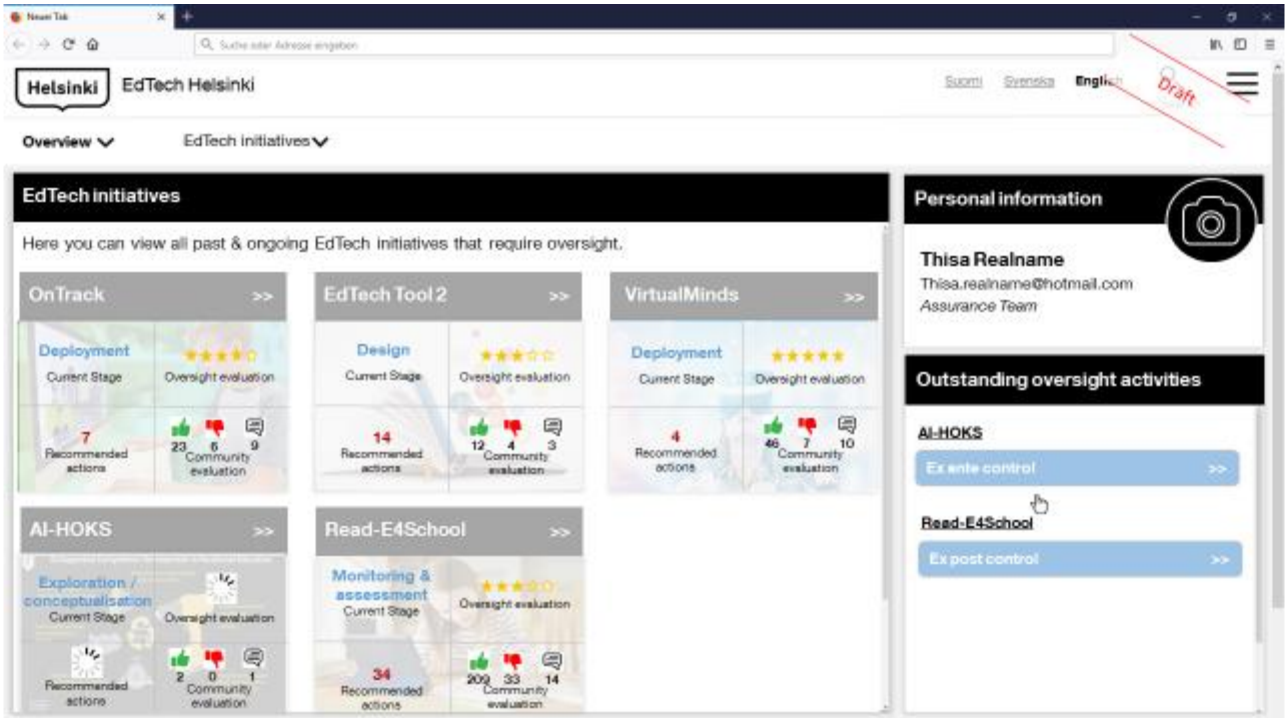


Figure 42: Overview page for oversight team members

On the overview page (Figure 42 & Figure 41), oversight team members can view all past & ongoing EdTech initiatives, their own personal information, as well as a list of outstanding oversight activities. Once an EdTech launcher has published a report (e.g., Figure 28), the oversight team members will be informed about this by receiving a new listing of the according oversight activity in the box "Outstanding oversight activities" (Figure 42 & Figure 41). For example, since the EdTech launcher of the initiative AI-HOKS has finished and documented the stage "Exploration / conceptualization", so that "Ex ante control" is now required by the oversight team members. Similarly, the EdTech launcher of the initiative Read-E4School has finished and documented the stage "Monitoring & assessment", so that "Ex post control" is now required by the oversight team members.

By clicking on a particular initiative, the oversight team member will be redirected to detailed information about this particular initiative (Figure 43).

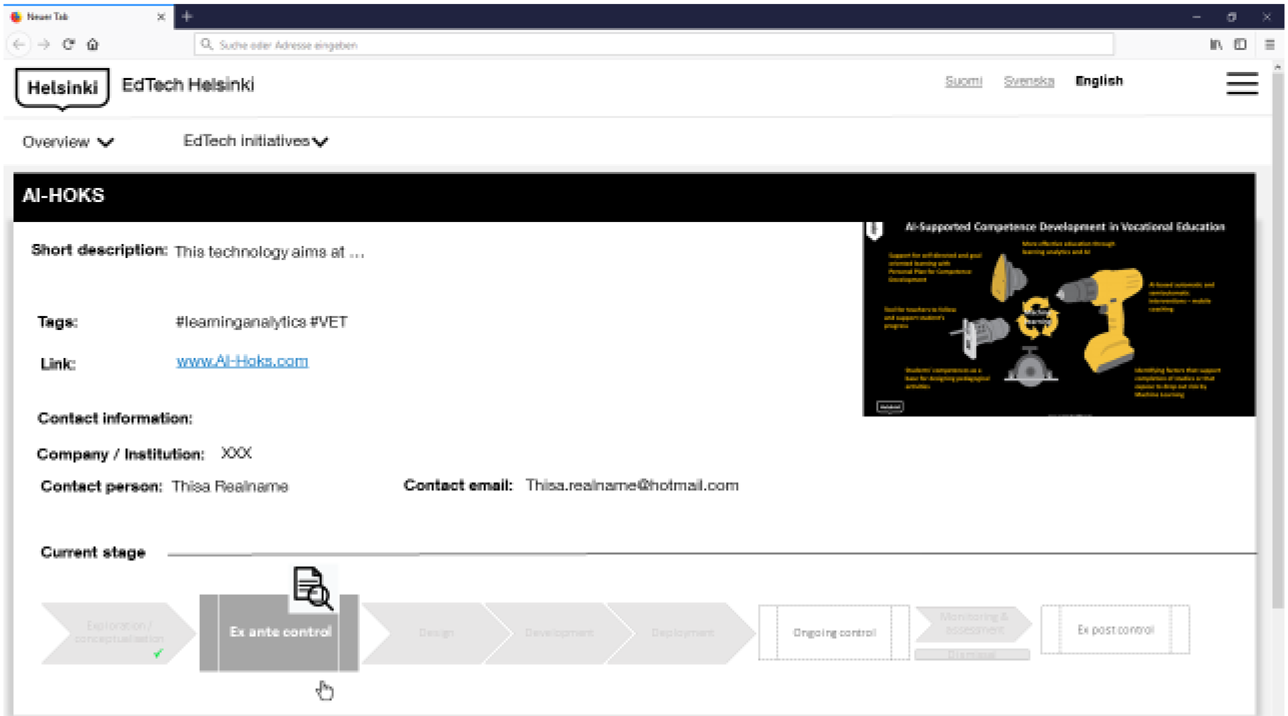
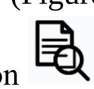


Figure 43: Detailed information about particular EdTech initiative AI-HOKS (for oversight team members)

Similar to the website flow of the community users (Figure 32), the detailed information that will be displayed here is the information that the responsible EdTech launcher had filled in when setting up the initiative (Figure 15).

In addition, the oversight team member can view the current stage of the particular initiative (Figure 43). For example, here the EdTech launcher of the AI-HOKS initiative has finished the “Exploration / conceptualization” stage and published an according report “Ex ante report” (Figure 28) so that “Ex ante control” by the oversight team can now be conducted. By clicking on , the oversight team member can view the report of the EdTech launcher and provide an oversight evaluation (Figure 44 & Figure 45).

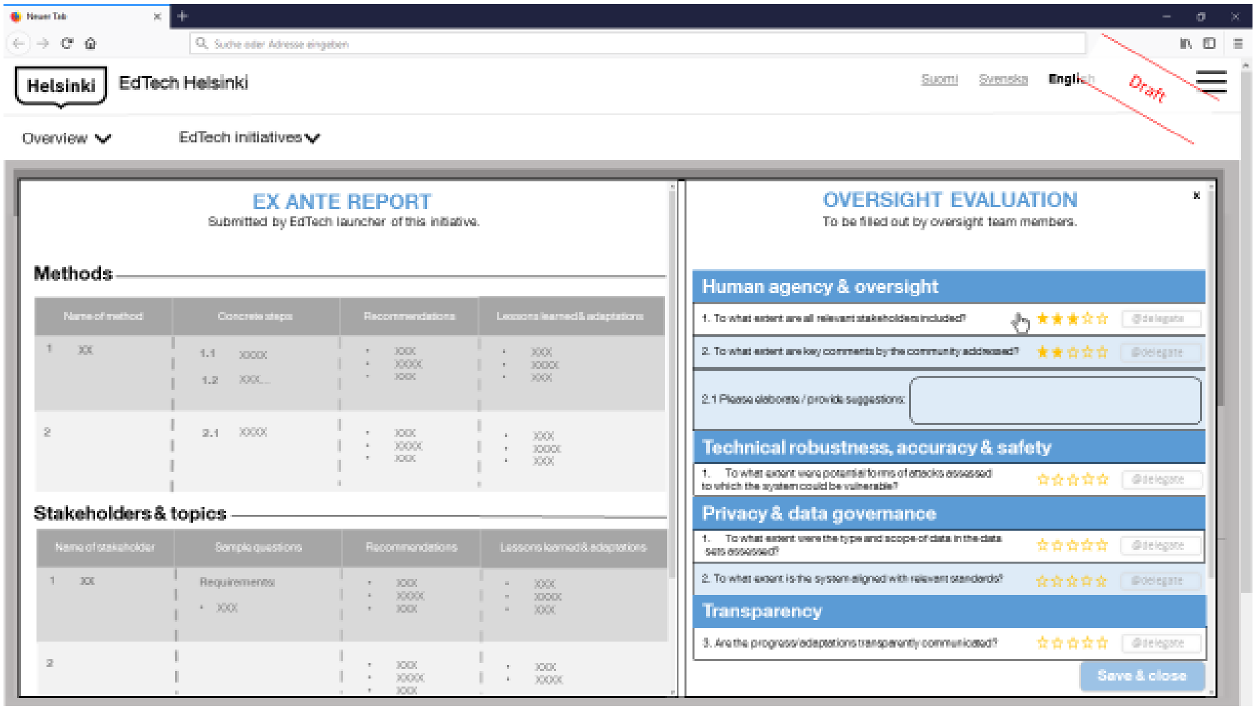


Figure 44: Ex ante control - Oversight evaluation (1)

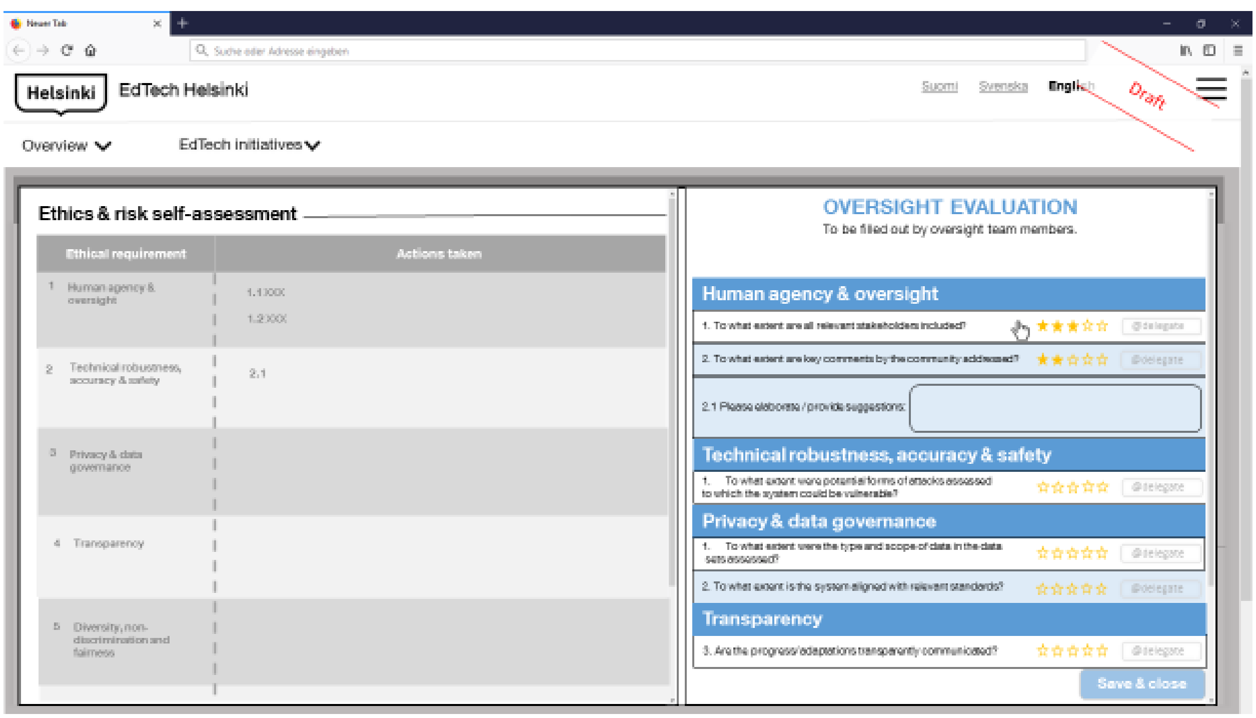


Figure 45: Ex ante control - Oversight evaluation (2)

When starting the "Oversight evaluation", for example, of the ex ante report (Figure 44 & Figure 45), oversight team members can:

- on the left hand side, view the report that was published by the EdTech launcher (Figure 28)
- on the right hand side, fill out a survey by evaluating to what extent particular topics / questions* are sufficiently addressed and if not sufficiently addressed, provide suggestions in a text box. These ratings and suggestions will be transferred back to EdTech launchers as well as aggregated in the "oversight rating" and "recommended actions" (e.g., Figure 22, Figure 31 & Figure 42).

The ex ante control and the ongoing control are equal in terms of the content that needs to be assessed by the oversight team. However, the ex post control differs in terms of content, as illustrated in the following.

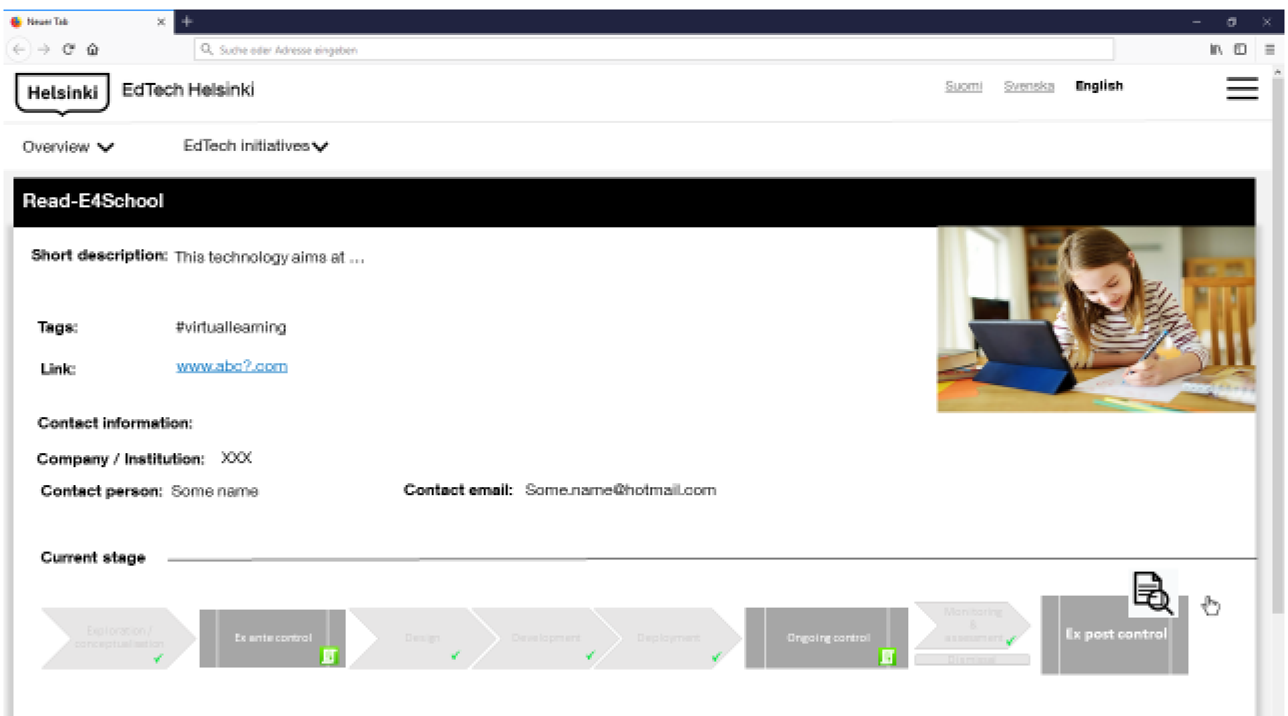


Figure 46: Detailed information about particular EdTech initiative Read-E4School (for oversight team members)

For example, the EdTech launcher of the Read-E4School initiative has finished the "Monitoring & assessment" stage and published a corresponding report "Ex post report" (Figure 30), so that "Ex post control" by the oversight team can now be conducted.

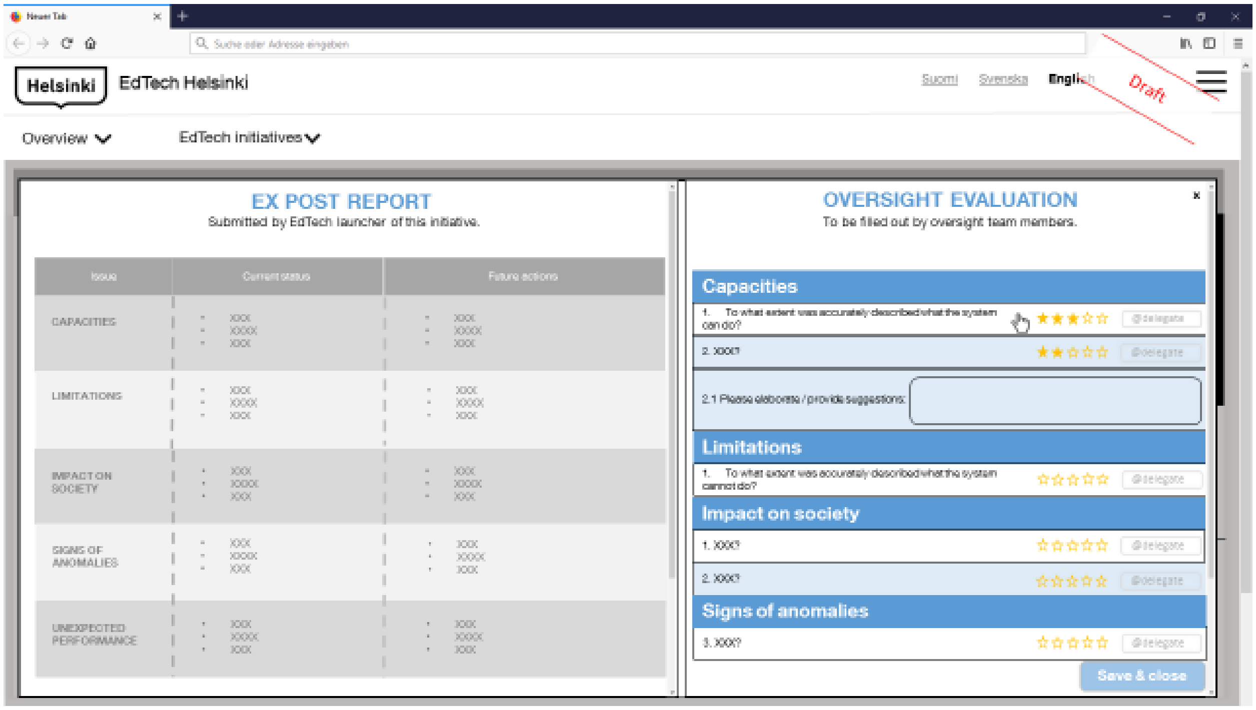


Figure 47: Ex post control – Oversight evaluation

When starting the "Oversight evaluation", for example, of the ex post report (Figure 47), oversight team members can:

- on the left hand side, view the report that was published by the EdTech launcher (Figure 30)
- on the right hand side, fill out a survey by evaluating to what extent particular issues** are sufficiently addressed and if not sufficiently addressed, provide suggestions in a text box. These ratings and suggestions will be transferred back to EdTech launchers as well as aggregated in the "oversight rating" and "recommended actions" (e.g., Figure 22, Figure 31 & Figure 42).

Relevant back-end material:	• *For more information, see "Translational Models" as well as the Assessment list for Trustworthy Artificial Intelligence by the High-level Expert Group on AI (2019). • **For more information, see "Appendix I: AI Ethics systemic translational matrix".

3 Summary

In this chapter, we will provide a short summary of the key features that were proposed in chapter 2, suggest how these can be integrated into the current effort of the City of Helsinki, and state necessary next steps.

3.1 Recap of key features

Key feature	Explanation	Corresponding issue that is addressed / achieved
Information disclosure	The platform / website informs on the issue at hand (i.e. education technology / learning analytics, the importance of stakeholder participation) as well as on all ongoing EdTech innovations throughout the entire life cycle including evaluations that are conducted by the oversight team.	● Transparency ● Human oversight ● Digital literacy
Overview pages	For EdTech launchers and the oversight team, this feature allows a concise summary of the ongoing (own) initiatives that allows improved monitoring of current stages, stakeholder evaluations, and pending to-dos.	● Transparency ● Human Oversight
Setting up initiatives & composing participatory toolkits	The platform / website provides EdTech launchers with an approach of how to structure the technological life cycle, as well as for each stage, how to set up a participatory toolkit (including the provision of concrete steps, sample questions and recommendations).	● Guidance ● Participation ● Human agency
Submitting ex ante / ongoing / ex post reports	For EdTech launchers, this feature provides prefabricated formats of report submissions. Thereby, EdTech launchers are informed on which issues they need to pay close attention to and report on.	● Efficiency ● Assurance ● Accountability ● Technical robustness, accuracy & safety
Co-design / ad-hoc comments	This feature allows the community to provide feedback and thereby shape the participatory toolkit of EdTech launchers in real-time. Furthermore, the community can identify participatory opportunities and sign-up for these.	● Transparency ● Human oversight ● Inclusion & Participation ● Assurance through collaboration
Ex post feedback	This feature allows the community to provide comments on the implemented EdTech initiatives (i.e., the utilized participatory toolkits as well as the reports published by the EdTech launcher).	● Transparency ● Human oversight ● Inclusion & Participation

		• Assurance through collaboration
Ex ante / ongoing / ex post control	For the oversight team, this feature provides prefabricated formats / surveys for overseeing the activities and submitted reports by EdTech launchers.	• Efficiency • Assurance • Accountability
Information gathering	All information concerning the content of the platform that is entered by the community and EdTech launchers should be transferred to the host of the platform, assessed and – if rational – picked up by the host by integrating it to the back end. This way, over time everything that will be displayed on the platform (e.g. proposed topics to be discussed or displayed when composing a participatory toolkit) will become more holistic.	• Continuous development, improvement & learning

3.2 Suggestions for integration

With websites such as the City of Helsinki AI Register (https://ai.hel.fi/en/ai-register/) and Design Helsinki (https://design.hel.fi/en/) or with platforms such as the one developed by Saidot (https://www.saidot.ai/), the City of Helsinki has already originated impressive initiatives to push for increased participation, as well as the democratic and ethical design of upcoming technologies. Therefore, the City of Helsinki has existing information infrastructure that could support and enable the deployment of our proposed type of multi-functional participation and accountability-enhancing web tool.

Building on Helsinki's – to some extent – separate initiatives, the proposed platform / website *EdTech Helsinki* should serve as an example for the development / progression of future participatory websites (e.g., some of the features proposed here can be easily transferred). In particular, the proposed platform / website *EdTech Helsinki* aims at showcasing how to better merge the strengths of each initiative. In particular, we propose that efforts such as Design Helsinki be combined with efforts such as the tool of Saidot into one single platform with varying interfaces / logins for different user types. Furthermore, this proposed platform should be linked to the City of Helsinki AI Register. For example, the information that is filled in for any EdTech initiative in the "basic information" form (Figure 15) could be directly transferred to the City of Helsinki AI Register. This way, the information is not only accessible via *EdTech Helsinki* and the City of Helsinki AI Register has the potential of becoming ever more holistic.

3.3 A few next steps

To ensure successful implementation of this proposed platform / website, the following steps may be necessary in the future:

1. Assess available / needed resources to further develop and implement this proposed platform.
2. Take a closer look at the stated limitations / things to consider and investigate potential solutions.
3. Check to what extent the proposed platform / website is in line with existing (data protection) regulations.
4. Develop the back end of this proposed platform. In particular, based on this Policy Paper, set up an excel / document with all relevant information (sample questions, recommendations, methods etc.) that will be displayed at the front end of this platform.

We wish you good luck with implementing this platform!

8 References

Ackoff, R. (1971). Towards a System of Systems Concepts. Management Science, 17(11), 661-671. Retrieved August 11, 2021, from http://www.jstor.org/stable/2629308

Ahonen, P., & Erkkilä, T. (2020). Transparency in algorithmic decision-making: Ideational tensions and conceptual shifts in Finland. *Information Polity*, *25*(4), 419–432. https://doi.org/10.3233/IP-200259

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004). Detecting Student Misuse of Intelligent Tutoring Systems. In J. C. Lester, R. M. Vicari, & F. Paraguaçu (Eds.), *Intelligent Tutoring Systems* (Vol. 3220, pp. 531–540). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-30139-4_50

Benthall, S., & Haynes, B. D. (2019). Racial Categories in Machine Learning. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 289–298. https://doi.org/10.1145/3287560.3287575

Bovens, M. (2007). Analysing and Assessing Accountability: A Conceptual Framework. European Law Journal, 13(4), 447–468. https://doi-org.libproxy.tuni.fi/10.1111/j.1468-0386.2007.00378.x

Brayne, S. (2014). Surveillance and System Avoidance: Criminal Justice Contact and Institutional Attachment. *American Sociological Review*, *79*(3), 367–391. https://doi.org/10.1177/0003122414530398

Brereton, M., & Buur, J. (2008). New challenges for design participation in the era of ubiquitous computing. *CoDesign*, *4*(2), 101–113. https://doi.org/10.1080/15710880802098099

Bright, J., Ganesh, B., Seidelin, C., & Vogl, T. (2019). *Data Science for Local Government*. Oxford Internet Institute, University of Oxford. https://smartcities.oii.ox.ac.uk/data-science-for-local-government-report/

Brkan, M. (2019). Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond. International Journal of Law and Information Technology, 27(2), 91–121. https://doi.org/10.1093/ijlit/eay017

Brunet, M. (2019). Governance-as-practice for major public infrastructure projects: A case of multilevel project governing. *Projects, Organizations and Institutions*, *37*(2), 283–297. https://doi.org/10.1016/j.ijproman.2018.02.007

Brunet, M., & Aubry, M. (2018). The governance of major public infrastructure projects: The process of translation. *International Journal of Managing Projects in Business*, *11*(1), 80–103. https://doi.org/10.1108/IJMPB-08-2017-0095

Centre of Excellence for Information Sharing. (2018). The General Data Protection Regulation—An opportunity for change. Centre of Excellence for Information Sharing. http://informationsharing.org.uk/gdpr/

Chambers, R. (2009). So that the poor count more: using participatory methods for impact evaluation. Journal of Development Effectiveness, 1(3), 243–246. https://doi.org/10.1080/19439340903137199

Croy, M. J. (1989). Ethical issues concerning expert systems' applications in education. *AI & SOCIETY*, *3*(3), 209–219. https://doi.org/10.1007/BF01891431

Data Futures Partnership. (2017). A Path to Social Licence: Guidelines for Trusted Data Use. Data Futures Partnership. https://toi-aria.s3.amazonaws.com/documents/Summary-Guidelines.pdf

Davis, J., Atchley, A., Smitherman, H., Simon, H., & Tenhundfeld, N. (2020). Measuring Automation Bias and Complacency in an X-Ray Screening Task. In 2020 Systems and Information Engineering Design Symposium (SIEDS) (pp. 1–5). IEEE. https://doi.org/10.1109/SIEDS49339.2020.9106670

De Stefano, V. (2020). Algorithmic Bosses and What to Do About Them: Automation, Artificial Intelligence and Labour Protection. In D. Marino & M. A. Monaca (Eds.), *Economic and Policy Implications of Artificial Intelligence* (pp. 65–86). Springer International Publishing. https://doi.org/10.1007/978-3-030-45340-4_7

Derakhshan, R., Turner, R., & Mancini, M. (2019). Project governance and stakeholders: A literature review. *International Journal of Project Management*, *37*(1), 98–116. https://doi.org/10.1016/j.ijproman.2018.10.007

Di Maddaloni, F., & Davis, K. (2017). The influence of local community stakeholders in megaprojects: Rethinking their inclusiveness to improve project performance. *International Journal of Project Management*, *35*(8), 1537–1556. https://doi.org/10.1016/j.ijproman.2017.08.011

Doll, R. (1974). Public benefit and personal privacy: The problems of medical investigation in the community. *Proceedings of the Royal Society of Medicine*, *67*, 1281–1285.

Dreyfus, H. L. (1992). *What computers still can't do: A critique of artificial reason* (3rd ed.). MIT Press.

Dreyfus, H. L. (2007). Why Heideggerian AI Failed and How Fixing it Would Require Making it More Heideggerian. *Philosophical Psychology*, *20*(2), 247–268. https://doi.org/10.1080/09515080701239510

Dreyfus, H. L., & Dreyfus, S. E. (1991). Making a Mind Versus Modelling the Brain: Artificial Intelligence Back at the Branchpoint. In M. Negrotti (Ed.), *Understanding the Artificial: On the*

Future Shape of Artificial Intelligence (pp. 33–54). Springer London. https://doi.org/10.1007/978-1-4471-1776-6_3

Dreyfus, H. L., Dreyfus, S. E., & Athanasiou, T. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. Basil Blackwell.

Dwyer, T. A. (1974). Heuristic strategies for using computers to enrich education. *International Journal of Man-Machine Studies*, *6*(2), 137–154. https://doi.org/10.1016/S0020-7373(74)80001-5

Ernst, A., Biß, K. H., Shamon, H., Schumann, D., & Heinrichs, H. U. (2018). Benefits and challenges of participatory methods in qualitative energy scenario development. Technological Forecasting and Social Change, 127, 245–257. https://doi.org/10.1016/j.techfore.2017.09.026

European Commission. Directorate General for Communications Networks, Content and Technology. (2020). The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment: Publications Office. https://doi.org/10.2759/002360Ferrario, M. A., Simm, W., Newman, P., Forshaw, S., & Whittle, J. (2014). Software Engineering for "Social Good": Integrating Action Research, Participatory Design, and Agile Development. *Companion Proceedings of the 36th International Conference on Software Engineering*, 520–523. https://doi.org/10.1145/2591062.2591121

Fadel, C., Bialik, M., & Holmes, W. (2019). Artificial Intelligence in Education. Promise and Implications for Teaching and Learning: Center for Curriculum Redesign. Retrieved from https://www.researchgate.net/publication/332180327_Artificial_Intelligence_in_Education_Promise_and_Implications_for_Teaching_and_Learning

Fink, M. (2021). The EU Artificial Intelligence Act and Access to Justice. EU Law Live.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *SSRN Electronic Journal*. Advance online publication. https://doi.org/10.2139/ssrn.3518482

Fukuda-Parr, S., & Gibbons, E. (2021). Emerging Consensus on 'Ethical AI': Human Rights Critique of Stakeholder Guidelines. Global Policy. Advance online publication. https://doi.org/10.1111/1758-5899.12965

Gacutan, J., & Selvadurai, N. (2020). A statutory right to explanation for decisions generated using artificial intelligence. International Journal of Law and Information Technology, 28(3), 193–216. https://doi.org/10.1093/ijlit/eaaa016

Giannouchos, T. V., Ferdinand, A. O., Ilangovan, G., Ragan, E., Nowell, W. B., Kum, H.-C., & Schmit, C. D. (2021). Identifying and prioritizing benefits and risks of using privacy-enhancing software through participatory design: A nominal group technique study with patients living with chronic conditions. Journal of the American Medical Informatics Association: JAMIA, 28(8), 1746–1755. https://doi.org/10.1093/jamia/ocab073

Gill, H. (2008). From vision to reality: cyber-physical systems. In HCSS national workshop on new research directions for high confidence transportation CPS: automotive, aviation, and rail (pp. 18-20). USA: Austin.

Gotterbarn, D., Miller, K., & Rogerson, S. (1999). Computer society and ACM approve software engineering code of ethics. Computer, 32, 84–88. https://doi.org/10.1109/MC.1999.796142

Gualdi, F., & Cordella, A. (2021). Artificial Intelligence and Decision-Making: The question of Accountability. Bui, Tung X., (Ed.) Proceedings of the 54th Annual Hawaii International Conference on System Sciences, HICSS 2021. Proceedings of the Annual Hawaii International Conference on System Sciences, 2297–2306. https://doi.org/10.24251/HICSS.2021.281

Hadden, S. G. (1986). Intelligent Advisory Systems for Managing and Disseminating Information. *Public Administration Review*, *46*, 572–578. JSTOR. https://doi.org/10.2307/975579

Harrington, C., Erete, S., & Piper, A. M. (2019). Deconstructing Community-Based Collaborative Design. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1–25. https://doi.org/10.1145/3359318

Head, S. (2005). *The New Ruthless Economy: Work and Power in the Digital Age*. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195179835.001.0001/acprof-9780195179835

Heimer, C. A. (2001). Cases and Biographies: An Essay on Routinization and the Nature of Comparison. *Annual Review of Sociology*, *27*(1), 47–76. https://doi.org/10.1146/annurev.soc.27.1.47

High-Level Expert Group on AI (2019). *Ethics guidelines for trustworthy AI*. European Commission.

Hood, C. (2007). What happens when transparency meets blame-avoidance? *Public Management Review*, *9*(2), 191–210. https://doi.org/10.1080/14719030701340275

Hurley, M. W., & Wallace, W. A. (1986). Expert Systems as Decision Aids for Public Managers: An Assessment of the Technology and Prototyping as a Design Strategy. *Public Administration Review*, *46*, 563–571. JSTOR. https://doi.org/10.2307/975578

Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. Educational Technology Research & Development, 64(5), 923–938. https://doi-org.libproxy.tuni.fi/10.1007/s11423-016-9477-y

Ingrams, A. (2020). A machine learning approach to open public comments for policymaking. *Information Polity*, *25*(4), 433–448. https://doi.org/10.3233/IP-200256

Jakobsen, M., Petersen, N. B. G., & Laumann, T. V. (2018). Acceptance or Disapproval: Performance Information in the Eyes of Public Frontline Employees. *Journal of Public Administration Research and Theory*, *29*(1), 101–117. https://doi.org/10.1093/jopart/muy035

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, 101493. https://doi.org/10.1016/j.giq.2020.101493

Jørgensen, K. (2016). The media go-along: Researching mobilities with media at hand. *MedieKultur: Journal of Media and Communication Research*, *32*(60). https://doi.org/10.7146/mediekultur.v32i60.22429

Kenny, S. (2008, May 1). An Introduction to Privacy Enhancing Technologies. The International Association of Privacy Professionals (IAPP). https://iapp.org/news/a/2008-05-introduction-to-privacy-enhancing-technologies/

Krafft, P. M., Young, M., Katell, M., Huang, K., & Bugingo, G. (2020). Defining AI in Policy versus Practice. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 72–78. https://doi.org/10.1145/3375627.3375835

Lang, C., Siemens, G., Wise, A., & Gasevic, D. (2017). Handbook of Learning Analytics: Society for Learning Analytics Research (SoLAR). https://doi.org/10.18608/hla17

Lassnigg, L. (2012). 'Lost in translation': Learning outcomes and the governance of education. *Journal of Education and Work*, *25*(3), 299–330. https://doi.org/10.1080/13639080.2012.687573

Lee, E. A. (2007). Computing foundations and practice for cyber-physical systems: A preliminary report. University of California, Berkeley, Tech. Rep. UCB/EECS-2007-72, 21.

Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, *20*(3), 881–900. https://doi.org/10.1177/1461444816675438

Malinverni, L., Schaper, M.-M., & Pares, N. (2019). Multimodal methodological approach for participatory design of Full-Body Interaction Learning Environments. *Qualitative Research*, *19*(1), 71–89. https://doi.org/10.1177/1468794118773299

Miailhe, N., Alanoca, S., Bora, A., Lannquist, Y., Arohi, J., Forer, G., . . . Falk, B. (2020). *Bridging AI's trust gaps: Aligning policymakers and companies*. Retrieved from EYQ (Ernst & Young's Think Tank) and The Future Society website: https://thefuturesociety.org/wp-content/uploads/2020/07/tfs-bridging-ais-trust-gaps-report.pdf

Miao, F., Holmes, W., Huang, R., Zhang, H. (2021). AI and Education: Guidance for Policy-Makers, UNESCO, ISBN 978-92-3-100447-6.

Mittelstadt, B. D., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 279–288.

Mohammed, P. S., & ‘Nell’ Watson, E. (2019). Towards Inclusive Education in the Age of Artificial Intelligence: Perspectives, Challenges, and Opportunities. In J. Knox, Y. Wang, & M. Gallagher (Eds.), *Artificial Intelligence and Inclusive Education* (pp. 17–37). Springer Singapore. https://doi.org/10.1007/978-981-13-8161-4_2

Mumtaz, S. (2000). Factors affecting teachers’ use of information and communications technology: A review of the literature. *Journal of Information Technology for Teacher Education*, *9*(3), 319–342. https://doi.org/10.1080/14759390000200096

Nevaranta, M., Lempinen, K., & Kaila, E. T. (2020). Students' Perceptions about Data Safety and Ethics in Learning Analytics. 1613-0073. Retrieved from https://helda.helsinki.fi/bitstream/handle/10138/325523/FP_2.pdf?sequence=1&isAllowed=y

O’Brien, L., Marzano, M., & White, R. M. (2013). ‘Participatory interdisciplinarity’: Towards the integration of disciplinary diversity with stakeholder engagement for new models of knowledge production. *Science and Public Policy*, *40*(1), 51–61. https://doi.org/10.1093/scipol/scs120

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. British Journal of Educational Technology,. doi:10.1111/bjet.12152.

Parkin, L., & Paul, C. (2011). Public good, personal privacy: A citizens’ deliberation about using medical information for pharmacoepidemiological research. *Journal of Epidemiology and Community Health*, *65*(2), 150–156. https://doi.org/10.1136/jech.2009.097436

Parton, N. (2006). Changes in the Form of Knowledge in Social Work: From the ‘Social’ to the ‘Informational’? *The British Journal of Social Work*, *38*(2), 253–269. https://doi.org/10.1093/bjsw/bcl337

Prinsloo, P., & Slade, S. (2014). Student privacy and institutional accountability in an age of surveillance. 10.1007/978-94-6209-794-0_12.

Prinsloo, P., & Slade, S. (2017). Ethics and learning analytics: Charting the (un) charted. SoLAR.

Reay, T., & Hinings, C. R. (2009). Managing the Rivalry of Competing Institutional Logics. *Organization Studies*, *30*(6), 629–652. https://doi.org/10.1177/0170840609104803

Sanders, E. B.-N., Brandt, E., & Binder, T. (2010). A Framework for Organizing the Tools and Techniques of Participatory Design. *Proceedings of the 11th Biennial Participatory Design Conference*, 195–198. https://doi.org/10.1145/1900441.1900476

Seale, J. (2010). Doing student voice work in higher education: an exploration of the value of participatory methods. British Educational Research Journal, Vol. 36, No. 6, pp. 995–1015

Shadowen, N., Lodato, T., & Loi, D. (2020). Participatory Governance in Smart Cities: Future Scenarios and Opportunities. In N. A. Streitz & S. Konomi (Eds.), LNCS Sublibrary: SL3 - Information Systems and Applications, incl. Internet/Web, and HCI: Vol. 12203. Distributed, ambient and pervasive interactions: 8th International Conference, DAPI 2020, held as part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, proceedings / Norbert Streitz, Shin'ichi Konomi (eds.) (Vol. 12203, pp. 443–463). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-030-50344-4_3

Shalini, & Tewari, A. (2020). Sustainable Education in India through Artificial Intelligence: Challenges and Opportunities. In 12th ACM Conference on Web Science (pp. 41–47). New York, NY, USA: ACM. https://doi.org/10.1145/3394332.3402828

Shneiderman B. (2020). Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems. ACM Trans. Interact. Intell. Syst. 10, 4, Article 26 (December 2020), 31 pages. DOI:https://doi.org/10.1145/3419764

Singapore Digital (SG:D). (2020a). *Compendium of Use Cases: Practical Illustrations of the Model AI Governance Framework—Volume 2*. Info-communications Media Development Authority (IMDA) and Personal Data Protection Commission (PDPC). https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf

Singapore Digital (SG:D). (2020b). *Model Artificial Intelligence Governance Framework: Second Edition*. Info-communications Media Development Authority (IMDA) and Personal Data Protection Commission (PDPC). https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf

Slade, Sharon & Tait, Alan. (2019). Global guidelines: Ethics in Learning Analytics.

Snow, T. (2021). From satisficing to artificing: The evolution of administrative decision-making in the age of the algorithm. *Data & Policy*, *3*, e3. Cambridge Core. https://doi.org/10.1017/dap.2020.25

Taylor, F. W. (1911). *The principles of scientific management*. Harper.

Thomson, A. J., & Schmoldt, D. L. (2001). Ethics in computer software design and development. Computers and Electronics in Agriculture, 30(1–3), 85–102. https://doi.org/10.1016/S0168-1699(00)00158-7

Unceta I., Nin J., Pujol O. (2020) Risk mitigation in algorithmic accountability: The role of machine learning copies. PLOS ONE 15(11): e0241286. https://doi.org/10.1371/journal.pone.0241286

UNESCO Institute for Lifelong Learning (2019). 4th global report on adult learning and education: Leave no one behind: participation, equity and inclusion.

Verdiesen, I., Santoni de Sio, F., & Dignum, V. (2021). Accountability and Control Over Autonomous Weapon Systems: A Framework for Comprehensive Human Oversight. *Minds and Machines*, *31*(1), 137–163. https://doi.org/10.1007/s11023-020-09532-9

Viale Pereira, G., Cunha, M. A., Lampoltshammer, T. J., Parycek, P., & Testa, M. G. (2017). Increasing collaboration and participation in smart city governance: A cross-case analysis of smart city initiatives. *Information Technology for Development*, *23*(3), 526–553. https://doi.org/10.1080/02681102.2017.1353946

Vincent-Lancrin, S. and R. van der Vlies (2020), "Trustworthy artificial intelligence (AI) in education: Promises and challenges", OECD Education Working Papers, No. 218, OECD Publishing, Paris, https://doi.org/10.1787/a6c90fa9-en.

Vogl, T. M. (2021). *Artificial Intelligence in Local Government: Enabling Artificial Intelligence for Good Governance in UK Local Authorities* (No. 2; p. 18pp). Oxford Commission on AI & Good Governance. https://oxcaigg.oii.ox.ac.uk/wp-content/uploads/sites/124/2021/04/Local-Government-report.pdf

Walonoski, J. A., & Heffernan, N. T. (2006). Detection and Analysis of Off-Task Gaming Behavior in Intelligent Tutoring Systems. In M. Ikeda, K. D. Ashley, & T.-W. Chan (Eds.), *Intelligent Tutoring Systems* (Vol. 4053, pp. 382–391). Springer Berlin Heidelberg. https://doi.org/10.1007/11774303_38

Watson, D. (2020). *Conceptual Challenges for Interpretable Machine Learning* (SSRN Scholarly Paper ID 3668444). Social Science Research Network. https://doi.org/10.2139/ssrn.3668444

Weld, D. S., & Bansal, G. (2019). The challenge of crafting intelligible intelligence. Communications of the ACM, 62(6), 70–79. https://doi.org/10.1145/3282486

Weyns, D. (2020). Towards a code of ethics for autonomous and self-adaptive systems. Proceedings of the IEEE/ACM 15th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, 163–165. https://doi.org/10.1145/3387939.3391567

White, S., Wastell, D., Broadhurst, K., & Hall, C. (2010). When policy o'erleaps itself: The 'tragic tale' of the Integrated Children's System. *Critical Social Policy*, *30*(3), 405–429. https://doi.org/10.1177/0261018310367675

Whitehill, J., Williams, J. J., Lopez, G., Coleman, C. A., & Reich, J. (2015). Beyond Prediction: First Steps Toward Automatic Intervention in MOOC Student Stopout. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2611750

Buchanan, W., Imran, M., Pagliari, C., Pell, J., & Rimpiläinen, S. (2020). Use of Participatory Apps in Contact Tracing: Options and Implications for Public Health, Privacy and Trust. https://doi.org/10.17868/73197

Zhou, J., Chen, F., Berry, A., Reed, M., Zhang, S., & Savage, S. (2020). A Survey on Ethical Principles of AI and Implementations. *In 2020 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 3010–3017). IEEE. https://doi.org/10.1109/SSCI47803.2020.9308437

Zuboff, S. (1985). Automate/informate: The two faces of intelligent technology. *Organizational Dynamics, 14*(2), 5–18. https://doi.org/10.1016/0090-2616(85)90033-6

Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. Basic Books.