START Foundation: Coping with Bias and Fairness when Implementing and Using an AI System

1 Introduction

Welcome! START Foundation is excited to have you on board as a new team member focusing on the rising field of digital ethics. Congratulations on taking on this important role!

START Foundation has embarked on a challenging project to integrate an AI-based recommender system into its recently launched digital learning platform. The primary objective of this project is to ensure that the AI-based recommender system is implemented with the highest standards of ethical responsibility and societal benefit. That is why START Foundation hired you: In your new role at the START Foundation, you are tasked with conducting a thorough examination of potential ethical issues of this AI system, particularly regarding bias and fairness.

1.1 Why START Foundation Is Concerned About Potential Bias and Fairness Issues

Let’s briefly delve into the theoretical background to better understand the concerns of START Foundation: Advances in machine learning (ML) and the availability of big data have led to the widespread application of AI systems in virtually all areas of business and daily life (Mehrabi, Morstatter, Saxena, Lerman, & Galstyan, 2021). AI is being used for a myriad of tasks (Cearley, Burke, Searle, & Walker, 2016), first and foremost for decision-making and decision support. Numerous decisions which were previously made by humans alone are now being made or assisted by AI systems. Many of these decisions have high social impact, such as in education or health. However, AI systems have been criticized for being opaque: they fail to explain the causal relationship between inputs and outputs, making accountability and meaningful scrutiny of algorithmic decisions difficult (Burrell, 2016).Opacity poses a major challenge because some AI systems, implemented for their efficiency and effectiveness benefits, can deploy unnoticed or unchecked biases at scale (McKinsey Global Institute, 2019). As a result, AI systems can expose individuals, organizations, and society to risks they were not even aware of (see examples in Figure 1).

This is why START Foundation is so concerned. As the START Foundation’s AI-based recommender system is intended to help START scholars navigate their educational and engagement journeys, it will play a decisive role in shaping their biographies. Unnoticed or unchecked biases in the AI system could have a significant impact on the future of the scholars. Therefore, coping with bias and fairness in this AI system is crucial.
1.2 Expectations and Preparation for Your Task

Farid Bidardel—START Foundation’s Chief Executive Officer (CEO)—has outlined specific questions for you to address (see Section 7). These questions focus on the factors START Foundation should consider regarding bias and fairness in its AI-based recommender system. To prepare you for your task, we will first provide you with some general knowledge about digital learning platforms, AI-based recommender systems, and START Foundation (Section 2-4). After that, Farid reports on two key cornerstones of the project: the origins workshop of the AI system and his phone call with Mira, an expert in digital ethics (Section 5-6). We also advise you to complete an interactive learning module on bias and fairness in AI systems. Available at https://bias-and-fairness-in-ai-systems.de/en/home/, this module promises a comprehensive and practical understanding of these important topics.

Enjoy tackling your task in your new role!

<table>
<thead>
<tr>
<th>Section 2</th>
<th>Section 3</th>
<th>Section 4</th>
<th>Section 5</th>
<th>Section 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry Background</td>
<td>START Foundation Background</td>
<td>Digital Transformation of START Foundation</td>
<td>Farid’s Insight: The Origins of the AI System</td>
<td>Farid’s Insight: Initial Concerns About Bias and Fairness</td>
</tr>
<tr>
<td>Content</td>
<td>Content</td>
<td>Content</td>
<td>Content</td>
<td>Content</td>
</tr>
<tr>
<td>General knowledge about digital learning platforms and AI-based recommender systems</td>
<td>General knowledge about START Foundation</td>
<td>General knowledge about START Foundation’s digital vision</td>
<td>Farid reports about the origins workshop of the AI-based recommender system</td>
<td>Farid reports on his phone call with Mira, during which Mira explained the concepts of bias and fairness</td>
</tr>
<tr>
<td>Learning Outcome</td>
<td>Learning Outcome</td>
<td>Learning Outcome</td>
<td>Learning Outcome</td>
<td>Learning Outcome</td>
</tr>
<tr>
<td>Understand the reasons for integrating an AI-based recommender system into a digital learning platform</td>
<td>Understand the historical context, development and impact of START Foundation</td>
<td>Understand what START Foundation is trying to achieve with its digital learning platform and for whom</td>
<td>Understand how the AI-based recommender system should work and what it should look like</td>
<td>Understand the concepts of bias and fairness</td>
</tr>
</tbody>
</table>

Figure 2. Preparation for Your Task

2 Industry Background

2.1 Digital Learning Platforms

Digital platforms have dramatically transformed access to information and services across multiple industries (Gawer & Henderson, 2007). This transformation has resulted in more efficient and diverse user interactions (Grover & Lyytinen, 2022) and expanded access to previously constrained resources (Constantinides, Henfridsson, & Parker, 2018). For instance, in the retail sector, the shift from physical stores to e-commerce platforms such as Amazon has broadened product accessibility. In healthcare, telemedicine platforms have enabled patients to consult with doctors remotely, overcoming geographic and time constraints.

The evolution of digital learning platforms has significantly impacted education (Mehta, Chauhan, Gupta, & Jaiswal, 2021). Particularly noteworthy are Massive Open Online Course (MOOC) platforms such Coursera,
edX, and Udemy\(^1\), which have experienced exponential growth, especially during the COVID-19 pandemic. These platforms have played a pivotal role in democratizing education by making courses from top universities accessible to anyone with an Internet connection (Lohr, 2020). Particularly, individuals from less affluent backgrounds benefit who previously faced cost constraints accessing educational resources. These platforms are characterized by global reach, open access, and online delivery (McAuley, Stewart, Siemens, & Cormier, 2010). Thereby, they transcend the constraints of traditional classroom settings and fixed schedules, facilitating educational access for a diverse range of learners (Nascimento Cunha, Chuchu, & Maziriri, 2020). While technological advancements offer numerous benefits, they also present challenges in ensuring effective learning. One notable issue is the lack of personalized support, leading to difficulties in navigating the vast array of available courses. Consequently, dropout rates and overall dissatisfaction are higher (Eom, Wen, & Ashill, 2006; Hone & El Said, 2016). AI-based recommender systems have the potential to address this issue.

### 2.2 AI-based Recommender Systems

Recommender systems have gained widespread popularity across various domains, recommending items on Amazon, movies or videos on Netflix and YouTube, and connections on LinkedIn (Gorgoglione, Panniello, & Tuzhilin, 2019; Jugovac & Jannach, 2017). These systems analyze user preferences and available options to guide users towards relevant choices (Ricci, Rokach, & Shapira, 2015). Through advancements in ML\(^2\), AI has notably enhanced capabilities (Chen et al., 2022; Mu, 2018). AI-based recommender systems excel at discovering complex, non-linear user-item relationships based on various data sources such as contextual, textual, and visual information (Q. Zhang, Lu, & Jin, 2021; S. Zhang, Yao, Sun, & Tay, 2020). Recent advances in AI-based recommender systems address limitations of traditional systems, such as data sparsity and the cold-start problem (Kumar & Sharma, 2017; Q. Zhang et al., 2021). Thus, high recommendation quality can be achieved (S. Zhang et al., 2020).

In digital learning platforms, an AI-based recommender system could, for example, assist learners with programming skills in making informed decisions about learning new programming languages or expanding their skills in complementary areas. This AI system could effectively serve as a personal mentor, guiding learners through their educational and engagement journey.

### 3 START Foundation Background

“START Foundation is a response to the lack of equal opportunities in Germany and thus a contribution to strengthening democracy. The better the educational and networking opportunities for young people with a migratory background, the greater the chance that they will be able to contribute their perspectives everywhere: in the workplace, in politics, in civil society”.


Animated by the results of the 2000 PISA study, the Hertie Foundation initiated the START scholarship program in 2002 to support the academic success and integration of young people with a migratory background. Becoming the START Foundation in 2007, the scholarship program expanded its reach in 2015 to support the growing number of refugees in Germany. START Foundation intensified its support for newly arrived young people, streamlining their integration into the German education system, and enhancing their chances of successfully completing their schooling. By 2018, START Foundation reshaped its mission to

---


\(^2\) As a subset of AI, ML enables machines to find patterns in data without explicit rule programming and to make decisions and predictions based on this analysis. This is made possible by the increasing availability of big data and high computing power. ML applications typically become more accurate the more data they have available – without the need for additional programming. Information from el Naqa and Murphy (2015).
empower young people from diverse backgrounds to contribute positively to democracy and society, focusing on future opportunities rather than origins.

![START Foundation's Logo and START Scholars](image)

**Figure 3. START Foundation’s Logo and START Scholars**

Today, the START Foundation stands out in Germany as a nationwide, individualized scholarship program for young people with a migratory background. It has a dedicated team of 21 staff members, a community of 500 START scholars, over 3,000 alumni from 63 different countries of origin, and valuable partnerships with political, business, and civil society institutions. It offers a comprehensive three-year individualized local education and engagement program that runs alongside regular school activities. Scholars actively participate in courses, workshops, and seminars covering a wide range of topics, including democracy, civic engagement, and digitalization. At the same time, scholars receive one-on-one guidance and mentorship, with financial support available as needed.

Upon completing the START scholarship program, these young people emerge with a well-defined vision for their personal and societal futures, understanding of their abilities, and confidence to navigate and address societal challenges. Graduates continue their engagement through volunteer programs and opportunities for ongoing participation and lifelong learning within the START alumni network.

4 Digital Transformation of START Foundation

*Over the past 20 years, START Foundation has made a significant contribution to strengthening the self-confidence of young people with a migratory background and providing them with the basis for active participation in society. I am always impressed by the diversity and commitment of our current and former START scholars. Now we want to work on reaching even more young people, supported by new digital possibilities.*


In early 2022, Farid Bidardel became CEO of START Foundation with a clear and ambitious vision to transform START Foundation into a digitally-forward organization. A cornerstone of his vision was to establish a digital learning platform alongside the existing scholarship program.

Up to that point, START Foundation selected only 180 scholarship recipients annually due to capacity constraints. Given a pool of over a thousand applications per year, many young people had to be rejected. Farid’s innovative approach aimed to broaden access by providing all applicants, regardless of their scholarship status, access to digital-only courses (MOOCs) on the newly envisioned digital learning platform. This initiative was designed to foster personal development and educational growth for all members of the START community. To realize this vision, START Foundation rapidly digitized its systems and processes in 2022. The introduction of the “START Campus” digital learning platform marked a pivotal shift in the organization’s operations and community engagement.

Launched in February 2023, START Campus offered over 400 MOOCs tailored to the variety of interests and educational needs of the START community. These innovative courses, such as “Human Rights”,
“Learn German Fast”, and “Public Speaking and Debate” aimed to empower young people with the knowledge and skills necessary for active participation in society.

Through this digital initiative, START Foundation expanded access to education. Under Farid’s leadership, START Foundation aims to empower every young person with a migratory background to thrive and contribute to society.

5 Farid’s Insight: The Origins of the AI System

In September 2023, the START Foundation took an important step to improve its digital learning platform by seeking feedback from its users. The feedback collected revealed difficulties in platform navigation and finding suitable educational and engagement opportunities. These challenges were exacerbated by the platform’s rapid expansion, which brought a wealth of opportunities but also complicated the user experience. Thus, the risks of decreased engagement and increased dropout rates became apparent.

In response to these pressing concerns, Farid took proactive measures. He appointed Gregory Grund as START Foundation’s new Digital Manager. Together, they brainstormed how to improve the platform’s usability and overall user experience. With an initial idea in mind, Farid and Gregory organized a two-day workshop on September 18 and 19, 2023.

In the following, Farid gives a detailed insight into the process and content of the workshop.

5.1 First Day of the Workshop

Gregory started the workshop with earnestness and clarity:

Welcome everyone. At the heart of START Foundation is our unwavering commitment to personalized mentoring. This means ensuring that our community members have access to opportunities that match their aspirations and contribute to both their personal growth and their community engagement. Currently, navigating our digital platform to find suitable educational and engagement opportunities can be daunting for users. Our offerings are broad, diverse, and constantly evolving. Without our personal guidance, it becomes a serious challenge for users to find what they really need on the platform.

Pausing for a moment to gauge the room’s attention, Gregory continued with a tone of solution-driven optimism:

Farid and I believe the answer lies in developing a tool that simplifies this process for our users. Imagine a digital education and engagement advisor. Imagine a system that intuitively guides users to those courses on our platform, that are tailored to the users’ individual preferences, skills, and aspirations. I’m thinking of an AI-based recommender system – a “Skill Compass”. This system would provide personalized support by learning from users’ past behavior to predict their current needs and preferences. It’s like having a personal mentor but powered by AI.

The team’s response was immediate and enthusiastic. Inspired by the idea, they spent the first day of the workshop brainstorming and creating mock-ups. Figure 4 illustrates the results of their efforts.

Upon logging into the digital learning platform, users are immediately confronted with a variety of different educational and engagement opportunities on the home screen. A call-to-action button at the top served as a gateway to the Skill Compass, inviting users to explore their personal education and engagement journey (see Screen 1). Clicking on the Skill Compass button opens a new realm of possibilities. Users are first greeted with a brief introduction to the benefits of receiving personalized recommendations. They can then check and revise their profile information by setting and modifying their preferences for skills and interests (see Screen 2). The visual presentation of recommended courses is designed to be user-friendly. The platform prominently highlights one recommended course by default (see Screen 3).
5.2 Second Day of the Workshop

On the second day of the workshop, the focus shifted to a deep dive into the intricacies of the project. The team dedicated this day to refining the project plan. May 2024 was set as a target for the release of the Skill Compass as a fully operational feature on START Campus. This ambitious goal was budgeted at an estimated cost of 250,000 Euros. To ensure clarity in execution, a detailed breakdown of roles, team members, and their specific responsibilities was outlined (see Table 1).

Table 1. The Project Team

<table>
<thead>
<tr>
<th>Roles</th>
<th>Responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project lead</td>
<td># Project management and coordination&lt;br&gt;# Ensuring the project's progress&lt;br&gt;# Conceptualization and evaluation of user testing</td>
</tr>
<tr>
<td><em>Farid, Gregory</em></td>
<td></td>
</tr>
<tr>
<td>AI specialists</td>
<td># Development of the AI-based recommender system (backend)&lt;br&gt;# Frontend development incl. user interface and interaction concept&lt;br&gt;# Integration of the Skill Compass into the digital learning platform incl. operation and monitoring concept</td>
</tr>
<tr>
<td><em>Kilian, Mike</em></td>
<td></td>
</tr>
<tr>
<td>Development of the digital learning platform&lt;br&gt;<em>Amrei, Tatjana</em></td>
<td># Coordination and preparation of the educational and engagement opportunities on the digital learning platform</td>
</tr>
</tbody>
</table>

Furthermore, the team created a first draft of the list of data categories for the Skill Compass database (see Table 2). To tailor the Skill Compass effectively to the diverse needs of the young people, it is essential to collect comprehensive information about them. This data will enable the Skill Compass to accurately
align the vast amount of educational and engagement opportunities with individual preferences and pinpoint suitable options.

Table 2. Categories of Data Available to Generate Recommendations (Draft)

<table>
<thead>
<tr>
<th>Categories</th>
<th>Examples</th>
<th>Intended Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification data</td>
<td>First name, last name, age, gender</td>
<td># Individual and personalized communication on the platform</td>
</tr>
<tr>
<td>Geographic information</td>
<td>School currently attended, place(s) of residence</td>
<td># Recommend courses tailored to a young person’s region or culture</td>
</tr>
<tr>
<td>Group membership of platform users</td>
<td>Scholars, alumni, users without a scholarship</td>
<td># Depending on the group to which the users belong, there are individual offers for further education and (political) engagement</td>
</tr>
<tr>
<td>Data from specific user query</td>
<td>Click behavior, scrolling behavior</td>
<td># Generate recommendations tailored to individual needs and continuously improve the quality of the recommendations (e.g., if a young person selects the goals of “strengthening democracy” and “learning German”, the Skill Compass can create a personalized learning profile)</td>
</tr>
<tr>
<td>Interaction data with recommendations</td>
<td>Skills, personal goals, preferred courses</td>
<td># Makes it possible to improve the quality of the recommendations in the long term (e.g., clicking on a detailed course description indicates that the user is interested in the recommended course)</td>
</tr>
<tr>
<td>Information on motivation and goals from the application process (if available)</td>
<td>Letter of motivation</td>
<td># Application for a START scholarship requires information about personal motivation and goals # Interest profile</td>
</tr>
<tr>
<td>Language skills</td>
<td>Mother tongue, foreign languages</td>
<td># Generate recommendations that match the user’s foreign language skills and current German language level # Such personalized recommendations can help young people stay motivated and make effective progress by enabling them to follow the learning material</td>
</tr>
<tr>
<td>Data on migration history</td>
<td>Country of origin, generation of immigrants</td>
<td># Targeted interactions between the young people and alumni can be promoted and intercultural competences can be strengthened</td>
</tr>
<tr>
<td>General user behavior</td>
<td>Successful logins, components accessed, search queries</td>
<td># Considerations of usage habits</td>
</tr>
<tr>
<td>Official application forms (if available)</td>
<td>Letter of recommendation</td>
<td># Letter of recommendation includes targeted questions on various aspects such as language skills, social/community involvement, and personal goal # Create a comparable profile of strengths and interests based on standardized and externally validated documents</td>
</tr>
</tbody>
</table>

6 Farid’s Insight: Initial Concerns About Bias and Fairness

The results of the two-day workshop left everyone euphoric and united in their vision for the Skill Compass. However, Farid remained acutely aware of the ethical implications surrounding digital technologies. He knew
it was imperative to address bias and fairness within the AI-based recommender system. Seeking expert
guidance, he reached out to his friend Mira, an expert in digital ethics.

In a phone call with Mira, Farid first shared his enthusiasm and concerns. Then, he asked for a brief
explanation of the concepts of bias and fairness. Mira, happy to share her knowledge, started with an
illustrative example:

Do you remember the COMPAS system used by the US Justice Department to predict criminal recidivism? It was found to be biased, particularly against black offenders. A study showed that black offenders were almost twice as likely as white offenders to be labeled higher risk, despite not reoffending. Similar biases have been observed in other areas, such as an AI system judging beauty pageant winners but showing bias against darker-skinned contestants. These biased predictions stem from hidden or neglected biases in AI systems.

She continued, delving into the complexities:

Bias can occur at any stage of the AI lifecycle, for example during data collection. When the underlying training data contains biases, the AI system trained on it will learn and reflect these biases in its predictions. For instance, AI systems for image classification are often trained on ImageNet, a set of more than 14 million labeled images. However, some groups are overrepresented while others are underrepresented in this data set. More than a third of ImageNet data originates from the US, although the US comprises less than five percent of the world’s population. China and India together contribute only three percent of ImageNet data, although they represent more than 36 percent of the world’s population. This lack of geodiversity has led to a proven bias towards Western cultures. An AI system trained on ImageNet classified a photo of a traditional US bride dressed in white as “bride”, “dress”, “woman”, and “wedding”, but classified a photo of an Indian bride as “performance art” and “costume”. This kind of bias is also known as representation bias.

Mira pointed out:

But bias doesn’t just stem from the underlying training data. Many AI systems receive feedback
as they make predictions. For example, a recommender system records which items the user
clicks on and how long they spend on those items. They treat this information as implicit signals
of relevance to refine the model. When a user clicks on the first item on a page of search results,
you have to ask yourself: Is it simply because it was the first item, or because it was actually the
most relevant to the user? Interpreting this feedback accurately is challenging. Additionally, bias
can arise from the way the AI system is modeled or evaluated. Algorithms may exhibit biased
behavior due to certain design choices, even if the data itself is unbiased. Bias can manifest in
various ways.

Mira then turned to the concept of fairness:

One approach to addressing concerns about bias in AI systems is to enforce group fairness.
Grouped fairness requires treating protected groups identified by sensitive attributes such as race,
gender, age, or national origin similarly to others. Adherence to fairness is necessary from both
ethical and legal perspectives.

Mira pointed out:

Unfairness can be illustrated through various examples, such as an AI system that approves or
rejects course registrations with unequal rates across genders. Formally defining fairness is
challenging. Fairness is a complex and multifaceted concept that depends on context and culture.

---

3 Information from Mehrabi et al. (2021).
5 Information from Teodorescu, Morse, Awwad, and Kane (2021); Sen, Dasgupta, and Gupta (2020).
There are over 20 mathematical definitions of fairness in the literature. “Demographic Parity” and “Equal Opportunity” are commonly used.

Mira concluded:

Addressing bias and fairness in your AI-based recommender systems is crucial. Let me say that these systems significantly influence users’ decisions and shape their experiences and outcomes. However, mitigating bias is challenging, as it can occur at any point in the AI lifecycle. First, potential biases have to be identified and an appropriate definition of fairness must be selected. Then, various mitigation approaches can be pursued, such as building a diverse development team, using explainable AI methods, or applying bias mitigation algorithms. Unfortunately, there is no “one-size-fits-all” approach. Rather, it depends on the situation. I can highly recommend an interactive learning module on bias and fairness in AI systems, where you can gain a deeper understanding and assess the issues in your concrete situation.

7 Your Task

Now that you have gathered all the necessary information for your task, you are well-prepared to address Farid’s questions. Your task will likely require further research on bias and fairness in AI-based recommender systems to ensure that your answers are comprehensive and well-informed. Therefore, as previously mentioned, we recommend the interactive learning module on bias and fairness in AI systems, available at https://bias-and-fairness-in-ai-systems.de/en/home/.

1. Discuss the relevance of unbiased and fair AI systems for business and society in general, as well as for START Foundation in particular.
2. What aspects of bias and fairness should START Foundation consider when developing the Skill Compass?
3. Should START Foundation embed the Skill Compass on its digital learning platform?

---

Information from Bellamy et al. (2018); Green and Hu (2020).
References


