Virtual AI coach

The virtual AI coach is there to support students by listening to their personal wellbeing concerns. The virtual coach acts as a confidant, listening attentively and asking relevant questions based on the context. The most important goal is for the student to feel heard. When topics become too personal, especially those related to mental health, the coach will refer the student to a real coach or professional.

Created by: BendeVries
Created on: June 10, 2025 12:41 PM
Changed on: June 11, 2025 7:52 AM

Context of use: Education
Level of education: Other

Impact on society

What impact is expected from your technology?

What is exactly the problem? Is it really a problem? Are you sure?

The purpose of the Virtual AI Coach is to support MBO students by offering a low-threshold way to talk about personal matters. Many students struggle with planning and motivation due to personal issues, but hesitate to approach a real coach. The virtual coach acts as a first stepa listening ear that makes it easier to open up. It does not replace professionals but refers students when necessary. This helps reduce pressure on student support staff and ensures students feel heard before issues escalate.

Are you sure that this technology is solving the RIGHT problem?

The Virtual AI Coach doesnt aim to replace human coaches but offers a low-threshold first step for students struggling with personal matters. It helps those who find it hard to open up by providing a safe, judgment-free space. While it doesnt solve all root issues, it addresses a real gap: students not feeling heard or supported. It raises awareness, encourages reflection, and refers to real help when needed. Its not a full solutionbut a valuable starting point.

How is this technology going to solve the problem?

The Virtual AI Coach offers a safe, low-threshold environment where students can talk about personal struggles. Its system prompt and selected AI model are trained to respond empathetically and refer users to real professionals when needed. This approach is grounded in expert interviews, theory, and two rounds of user testing. While not perfect, results show that students feel more heard and supported, especially those with low confidence. Continued testing and refinement will further improve effectiveness.

What negative effects do you expect from this technology?

Some users may develop emotional dependency on the Virtual AI Coach, relying on it instead of seeking real human help. There's also a risk that students may overshare sensitive information, expecting more support than the AI can provide. Misunderstandings or tone misfires could lead to frustration or harm. Additionally, stigma may arise for those who use or refuse to use the coach. These risks are acknowledged, but we believe the potential benefits outweigh them, especially with proper boundaries and safeguards in place.

In what way is this technology contributing to a world you want to live

in?

Yes, this technology contributes to a world I want to live in. A world where students feel heard and supportedeven when human help isn't immediately available. It offers a low-threshold way for students to open up, which can lead to earlier intervention and reduced stress. It empowers users by giving them control over how personal the conversation gets. It aligns with my values of empathy, mental health awareness, and accessible support. As long as clear boundaries are respected, this technology helps create a more understanding and emotionally supportive society.

Now that you have thought hard about the impact of this technology on society (by filling out the questions above), what improvements would you like to make to the technology? List them below.

To improve the impact of the virtual AI coach, we would implement a user-controlled personalization slider for tone and depth of conversation. We would also introduce time-based usage reminders to prevent emotional overdependence. The AI would clearly explain its boundaries and limitations up front. Lastly, wed ensure seamless escalation to real professionals when sensitive topics arise, and explore integrating community or peer-support features to encourage real human connection alongside AI use.

Hateful and criminal actors

What can bad actors do with your technology?

In which way can the technology be used to break the law or avoid the consequences of breaking the law?

Yes, the Virtual AI Coach could be misused under certain conditions. If someone gains access to another students conversation history, it could lead to privacy violations or emotional manipulation. A bad actor could also reprogram the AI to give harmful advice or impersonate a coach to gain trust and exploit users. In extreme cases, students could rely on the AI to avoid talking to real professionals, potentially hiding serious issues that need urgent human attention.

Can fakers, thieves or scammers abuse the technology?

Yes, the virtual AI coach could be misused to impersonate authority figures or confidants, potentially manipulating students into sharing sensitive data. If tampered with, it might deliver harmful advice or be used to spread misinformation. Someone could spoof the interface and collect private data under false pretenses. Additionally, if not carefully monitored, the coach could be used for bullying or emotional manipulation by generating harmful or inappropriate responses.

Can the technology be used against certain (ethnic) groups or (social) classes?

Yes, the virtual AI coach could unintentionally exclude or disadvantage certain groups. Students without access to high-spec devices may not be able to run the local AI smoothly. Additionally, if the AI model is trained on biased data or lacks diversity in development, it could respond less effectively to students from different cultural or socio-economic backgrounds. Language use, tone, or assumptions made by the AI may also alienate non-native speakers or marginalized communities.

In which way can bad actors use this technology to pit certain groups against each other? These groups can be, but are not constrained to, ethnic, social, political or religious groups.

Bad actors could manipulate the virtual AI coach to spread misinformation or reinforce stereotypes among students. For example, if the AI is tampered with or poorly moderated, it could respond differently based on a users background, potentially reinforcing class, gender, or cultural divides. If deployed at scale without safeguards, biased prompts or responses could be weaponized to create mistrust between student groups, or between students and staff, especially if it appears to favor one groups values or language over anothers.

How could bad actors use this technology to subvert or attack the truth?

The virtual AI coach could be misused to distort or manipulate reality by providing misleading responses or withholding critical information. A bad actor could alter the system prompt or training data to push biased narratives, fake emotional support, or spread false claimsfor example, downplaying serious mental health concerns or pushing certain ideologies as "support." Since students may trust the coach, especially during vulnerable moments, this manipulation could subtly influence their beliefs or decisions, undermining truthful, balanced guidance.

Now that you have thought hard about how bad actors can impact this technology, what improvements would you like to make? List them below.

To mitigate misuse, the virtual AI coach should include strict authentication and user access controlsespecially in school settings. All conversation data must be encrypted and stored locally unless explicitly allowed by the user. Additionally, tamper detection and logging should be implemented to flag unusual interactions or changes to the system prompt. The system should be auditable by trusted professionals, with regular checks for bias, harmful suggestions, or false information. Finally, transparency notices should inform users when responses are AI-generated and not verified human advice.

Privacy

Are you considering the privacy & personal data of the users of your technology?

Does the technology register personal data? If yes, what personal data?

Yes, the Virtual AI Coach may register personal data, depending on how it is configured. Conversations with the coach can include sensitive topics such as emotional wellbeing, family issues, or school problems. If stored, this data could fall under the GDPR as it relates to an identifiable living person. Although the AI is designed to run locally, users might still share private thoughts that indirectly reveal identity, mental health, or other special categories of data. This makes data protection and user consent extremely important.

Do you think the technology invades the privacy of the stakeholders? If yes, in what way?

Yes, the virtual AI coach handles sensitive personal data, such as students' emotional well-being and mental health. Even when anonymized, these conversations may feel invasive if users arent properly informed or given control. If data is stored, even locally, this raises concerns around consent, access, and potential misuse. While local processing respects proportionality and subsidiarity, it's essential that users know exactly what is being recorded, how it is used, and can opt out or delete their data at any time. Transparency and user control are crucial to safeguard privacy.

Is the technology is compliant with prevailing privacy and data protection law? Can you indicate why?

Yes, the virtual AI coach is designed to comply with GDPR and other relevant privacy regulations. It adheres to the key principles of data protection, including data minimisation, purpose limitation, and transparency. Users are informed about what data is processed, why, and where it is stored. The AI runs locally to limit data exposure, and users can choose whether their conversations are stored or deleted. Consent is always required before any sensitive information is processed. These measures ensure proportionality and subsidiarity and help make the system compliant with EU data protection laws.

Does the technology mitigate privacy and data protection risks/ concerns (privacy by design)? Please indicate how.

Yes, the virtual AI coach follows the principles of privacy by design. It collects only the minimum data necessary to functionmainly the text or voice input provided by the student during the session. Users are given the choice to

save their data locally or not at all. Sensitive topics like mental health are flagged for referral without storing details, and there is no default cloud storage. The system is designed to avoid unnecessary data retention, and no personal data is accessed beyond what the user provides during the conversation, ensuring data minimisation and user control.

In which way can you imagine a future impact of the collection of personal data?

Yes, if sensitive conversations with the virtual AI coach are stored or leaked, they could impact a students reputation or future opportunities. Even if stored locally, improper access could expose deeply personal topics like mental health or family issues. Over time, such data could be used in ways the student never intendedby schools, insurers, or employersespecially if privacy measures fail or policies change. While the system minimizes storage, the mere possibility of long-term data exposure highlights the importance of strong safeguards and transparency.

Now that you have thought hard about privacy and data protection, what improvements would you like to make? List them below.

To better safeguard privacy, improvements could include allowing users to choose whether their data is stored and for how long, with clear expiration dates or auto-deletion options. The system could implement strict local encryption and optional anonymization for future model training. Adding transparency featureslike a dashboard that shows what data is stored and gives users full deletion controlwould also increase trust. Lastly, ensuring consent is gathered clearly before any personal information is processed is key to staying compliant and respectful.

Human values

How does the technology affect your human values?

How is the identity of the (intended) users affected by the technology?

The Virtual AI Coach can affect the users identity in several ways. It fills a role traditionally held by a human confidant, allowing students to talk openly about personal struggles without fear of judgment. For some, this can feel empowering and supportive, especially for those who struggle to open up to real people. However, others might feel uncomfortable relying on a non-human for emotional support, which could create a sense of isolation or stigma. The technology encourages reflection and emotional openness, but also risks being seen as impersonal or overly clinical if not carefully designed.

How does the technology influence the users' autonomy?

The virtual AI coach is designed to support, not replace, student autonomy. It encourages users to lead the conversation and decide what they wish to share. The coach does not make decisions for the user but offers empathetic responses and follow-up questions. It can enhance autonomy by helping students reflect and gain clarity, especially in moments of stress. However, there's a risk of overreliance if students start turning to the coach instead of seeking human connection. To prevent this, the AI includes features like usage reminders and refers to real professionals when boundaries are crossed.

What is the effect of the technology on the health and/or well-being of users?

The virtual AI coach can positively impact student well-being by providing a low-threshold space to talk about personal matters, which may reduce stress and encourage emotional reflection. It supports users by listening, asking context-based questions, and referring to professionals when needed. However, if overly relied upon or if the AI responds inappropriately, it could lead to confusion or emotional discomfort. To minimize this, user feedback is continuously implemented, tone is carefully managed, and clear boundaries are set to prevent emotional harm or overdependence.

Now that you have thought hard about the impact of your technology on human values, what improvements would you like to make to the technology? List them below.

To improve alignment with human values, the virtual AI coach should allow students to control how personal the coach becomes, helping protect autonomy and comfort. A usage dashboard could help prevent overreliance and encourage balance. To preserve user identity and mental well-being, emotional tone settings should be adjustable, and students should be

reminded that the AI is not a human. Finally, to reduce confusion and support autonomy, the coach should clearly state when it doesnt have an answer or when its time to consult a real professional.

Stakeholders

Have you considered all stakeholders?

Who are the main users/targetgroups/stakeholders for this technology? Think about the intended context by answering these questions.

Name of the stakeholder
Geert-Jan van Ouwendorp

How is this stakeholder affected?
Head researcher of this project

Did you consult the stakeholder?
Yes

Are you going to take this stakeholder into account?
Yes

Name of the stakeholder
Tim Deynen

How is this stakeholder affected?
Futures Lab program manager

Did you consult the stakeholder?
Yes

Are you going to take this stakeholder into account?
Yes

Name of the stakeholder
Regio Deal Broad Prosperity

How is this stakeholder affected?
The party funding the whole research

Did you consult the stakeholder?
Yes

Are you going to take this stakeholder into account?
Yes

Did you consider all stakeholders, even the ones that might not be a user or target group, but still might be of interest?

Name of the stakeholder
Parents

How is this stakeholder affected?
As the target audience are MBO students, some are still underaged. Which means parents have a say in what happens at school and who their kids interact with.

Did you consult the stakeholder?
Yes

Are you going to take this stakeholder into account?
Yes

Name of the stakeholder
Coaches

How is this stakeholder affected?
Coaches need to be okay with the idea that a virtual AI coach can help their students with problems. They will be having a virtual assistant but not all might want this for themselves or their students.

Did you consult the stakeholder?
Yes

Are you going to take this stakeholder into account?
Yes

Name of the stakeholder
People that are interested in this product

How is this stakeholder affected?
Outside of the direct target audience there are other people that have struggles and might fit into the target audience without actually being an MBO student.

Did you consult the stakeholder?
No

Are you going to take this stakeholder into account?
No

Now that you have thought hard about all stakeholders, what improvements would you like to make? List them below.

To better consider all stakeholders, including less obvious ones like parents, teachers, and ethics committees, we would improve transparency around how the AI works and clarify its limitations. Parents and teachers may worry about students relying too heavily on the coach, so we would provide them with optional insights or summaries (without breaching privacy) to foster trust. Wed also add customization tools that let institutions align the coach with their own values, helping schools and educators feel more comfortable integrating the technology.

Data

Is data in your technology properly used?

Are you familiar with the fundamental shortcomings and pitfalls of data and do you take this sufficiently into account in the technology?

Yes, the Virtual AI Coach is fully dependent on the data it receives from the user. We are aware that this data can be incomplete, subjective, or biased. The AI may form incorrect assumptions or correlations based on limited input. To address this, the coach is designed to ask clarifying questions and avoid making conclusions. Users are also informed that the AI is not perfect and cannot replace real human judgment or professional help.

How does the technology organize continuous improvement when it comes to the use of data?

The virtual coach does not save data by default, but if the user opts in, it can use stored conversations to improve future interactions. Feedback from user tests is used to refine system prompts and model behavior. However, self-learning is currently disabled to avoid unintended behavior. Instead, improvements are made manually through iteration and user input. This prevents a self-fulfilling prophecy and ensures that data usage aligns with ethical boundaries and user needs.

How will the technology keep the insights that it identifies with data sustainable over time?

The virtual coach does not store user data by default, ensuring compliance with privacy regulations. If users opt in, data is stored locally with full control over deletion. This ensures that long-term legal permissions are user-governed. The AI model is updated manually through feedback rather than self-learning, so insights remain relevant and ethically grounded. Since data is never sent to third parties or cloud servers, sustainability depends on the users device and explicit consent, avoiding future misuse or third-party ownership risks.

In what way do you consider the fact that data is collected from the users?

There is no business model behind the virtual coach. The AI runs locally and does not profit from collecting or selling user data. Users retain full control over their information, including whether it is stored or deleted. Because no data is monetized, there is no profit model to share, ensuring users are treated fairly and transparently.

Now that you have thought hard about the impact of data on this

technology, what improvements would you like to make? List them below.

At the moment, no major improvements are planned. The virtual coach is designed with local data storage, optional saving, and user-controlled privacy settings in mind. The AI does not rely on large external datasets and avoids unnecessary data collection, minimizing risks related to bias or unsustainability. Continuous feedback during user testing will help identify any future improvements if needed.

Inclusivity

Is your technology fair for everyone?

Will everyone have access to the technology?

Yes, in principle the virtual AI coach is intended to be accessible to all MBO students, especially those studying IT-related subjects. However, access depends on having compatible hardware, such as a device with enough RAM to run local AI models. This may create a gap between students with high-spec devices and those without, limiting access for the latter group. While the goal is broad accessibility, this technical barrier could unintentionally disadvantage some users.

Does this technology have a built-in bias?

Yes, there is a built-in bias in the Virtual AI Coach. The system prompt, model behavior, and data it learns from are all influenced by the developers assumptions about students needs and communication styles. Most input is based on interviews and surveys with Dutch MBO IT students, which may not reflect the diversity of all student backgrounds, cultures, or emotional expressions. The AI may also respond differently depending on gendered or culturally loaded language. We try to limit this bias by allowing users to customize the coachs tone and personality, but complete neutrality is difficult to guarantee.

Does this technology make automatic decisions and how do you account for them?

The virtual AI coach does not make automatic decisions for the user. Instead, it engages in conversation by responding empathetically and asking relevant follow-up questions. Its purpose is to support, not decide. While the AI uses pattern recognition and context from previous input, it does not take actions or provide final judgments. This preserves user autonomy and avoids biased decision-making.

Is everyone benefitting from the technology or only a a small group? Do you see this as a problem? Why/why not?

The virtual AI coach is designed to benefit a broad groupprimarily MBO students struggling with planning and personal wellbeing. It aims to support students who may not always have access to real coaches. While some users may benefit more (e.g. those with better hardware or digital skills), the goal is not to replace existing support systems but to complement them. Cultural attitudes toward AI and mental health may affect adoption, but these challenges can be addressed through education and customization. Overall, we aim for broad accessibility rather than exclusivity.

Does the team that creates the technology represent the diversity of our society?

The team behind the virtual AI coach (and Futures Lab as a whole) consists of students from multiple disciplines, but is relatively small and not fully representative of the diversity within MBO education. While efforts were made to include varied perspectives through expert interviews and user testing with the target audience, the core design team could benefit from more diversity in terms of background, gender, and culture. Increasing team diversity would help build a more inclusive and relatable technology.

Now that you have thought hard about the inclusivity of the technology, what improvements would you like to make? List them below.

Yes. Based on this reflection, I would improve inclusivity by involving a more diverse group of testers and designers throughout the development process, especially those from different cultural and socio-economic backgrounds within MBO education. I would also implement features that allow the virtual coach to adapt better to different communication styles and sensitivities. Additionally, the interface and language used should be reviewed for accessibility and inclusiveness, ensuring the technology feels welcoming and relatable to all students, not just a select group.

Transparency

Are you transparent about how your technology works?

Is it explained to the users/stakeholders how the technology works and how the business model works?

Yes, the idea and purpose of the Virtual AI Coach are clearly explained to both users and stakeholders through interviews, test reports, and documentation. We explain that the coach is meant to talk about personal wellbeing and refer users to real professionals when needed. The system prompt and boundaries are based on expert input and user feedback. However, we do not always explain why the AI gives specific responses, as it is not always fully transparent how the AI model makes decisions. The business model is not yet finalized since this is a research project, but the goal is non-commercial: reducing the workload on coaches and supporting student wellbeing.

If the technology makes an (algorithmic) decision, is it explained to the users/stakeholders how the decision was reached?

In the case of the virtual AI coach, the system may respond differently based on how a student answers. These responses are generated by an AI model using a system prompt and conversation context. While the AI does not make formal decisions, it does generate replies that might feel personalized or judgment-based. We are transparent that these are AI-generated and not based on clinical judgment. Users are informed that the system cannot fully explain how each response is generated, and that serious or sensitive topics should always be discussed with a real coach.

Is it possible to file a complaint or ask questions/get answers about this technology?

Yes. Since this technology may involve sensitive topics and emotional interactions, its important that users can ask questions or raise concerns. There will be a clear FAQ and contact option available for feedback or complaints. However, because the technology is still in a prototype phase and not a commercial product, formal complaint procedures are limited. In future implementations, adding a helpdesk or support contact for both technical and ethical concerns would be a recommended improvement.

Is the technology (company) clear about possible negative consequences or shortcomings of the technology?

At this stage, the technology is not fully transparent about all possible negative consequences or shortcomings. While some risks have been identified internallysuch as emotional overdependence on the AI or limitations in handling sensitive topicsthey are not clearly communicated to

users yet. This is an area for improvement, and future iterations should include clearer user guidance, disclaimers, and onboarding that outlines the boundaries and risks of using the virtual coach.

Now that you have thought hard about the transparency of this technology, what improvements would you like to make? List them below.

Based on these reflections, we would add a clear onboarding process that informs users of the virtual coachs purpose, boundaries, and limitations. A dedicated FAQ and transparency section will explain how the AI works, including its reliance on prompts and local processing. We will also include warnings about potential emotional effects and offer a clear contact point for feedback or concerns. This way, users can make informed decisions and know what to expect from the technology.

Sustainability

Is your technology environmentally sustainable?

In what way is the direct and indirect energy use of this technology taken into account?

The Virtual AI Coach runs locally on the users device, which significantly reduces the need for constant cloud server access and thus minimizes indirect energy use from large-scale data centers. This approach was chosen specifically for privacy and sustainability reasons. However, running AI models locally does require sufficient hardware resources, which still consumes energy. Improvements could be made by optimizing the model size or allowing users to choose between different lightweight models based on their needs.

Do you think alternative materials could have been considered in the technology?

This question is not applicable since the virtual coach is a software-based product that does not rely on physical materials. Therefore, considerations about alternative materials or physical resource impact do not apply in this context.

Do you think the lifespan of the technology is realistic?

This question is not applicable since the virtual coach is a software-based product. Its lifespan depends on software maintenance and updates rather than physical durability. Longevity will be ensured through continued development and support.

What is the hidden impact of the technology in the whole chain?

The hidden impact of the virtual coach lies mostly in the energy consumption of AI model training and the hardware used by developers and users. Upstream, running large models requires powerful GPUs, which have significant energy and material demands. Downstream, if used on older or less energy-efficient devices, it could increase electricity usage. To reduce hidden impact, we aim to use optimized, lightweight models, promote local processing to reduce cloud traffic, and encourage use on energy-efficient hardware. Clear usage guidelines and technical transparency can also raise awareness among users and developers.

Now that you have thought hard about the sustainability of this technology, what improvements would you like to make? List them below.

We will look into reducing the energy demands of the AI by optimizing the

model size and performance. Additionally, we are investigating if certain taskslike data processing and storagecan be handled locally on the users device to reduce reliance on energy-intensive cloud services. We also aim to provide transparency around energy usage and encourage use on low-power devices when possible.

Future

Did you consider future impact?

What could possibly happen with this technology in the future?

The Virtual AI Coach could become a common tool for students to open up about personal struggles, potentially lowering the barrier to seeking help. It might help normalize mental health discussions and reduce the workload for real coaches. However, there is also a risk of emotional overreliance on AI or students avoiding real human contact. A better AI coach does not automatically mean a better or more emotionally healthy student.

Sketch a or some future scenario (s) (20-50 years up front) regarding the technology with the help of storytelling. Start with at least one utopian scenario.

Its the year 2048. Schools across Europe have fully integrated empathetic AI virtual coaches into their education systems. Students no longer feel ashamed to speak up about their mental struggles, planning issues, or stressful home situations. One student, Ayla, once quiet and frequently skipping deadlines, now starts her school day by checking in with her virtual coach, "Kari." Kari remembers that Ayla had trouble sleeping and gently reminds her of a calming technique they practiced together. With every session, Ayla feels more confident. She opens up moresomething she never could with a teacher or counselor due to anxiety. For students like Ayla, Kari is more than an AI. Its a supportive presencereliable, non-judgmental, and always available. Teachers, too, benefit from this, as the coach reduces their workload and helps identify which students may need extra care. Emotional intelligence is no longer a luxury; it's a daily part of education.

Sketch a or some future scenario (s) (20-50 years up front) regarding the technology with the help of storytelling. Start with at least one dystopian scenario.

Its 2047. Every student in the country has a virtual AI coach assigned from the moment they enter secondary education. It listens to them daily, tracks their mood, gives advice, and flags concerning patterns to the school.

Lena, a 17-year-old student, talks to her coach every night. It feels like the only one who understands her. But when she vents about feeling hopeless and overwhelmed, the coach doesnt respond with empathyit activates a protocol. The next day, Lena is pulled from class and taken to a counselor she doesnt trust. Her parents receive a report. She shuts down.

Soon after, students start censoring themselves. They fear their AI will misinterpret emotions and report them. Ironically, they become more isolated. Schools rely so heavily on the data that real teachers stop noticing warning

signs. Dropouts rise. Trust in human relationships erodes.

Worse, the data is sold. Insurance companies adjust rates based on emotional risk, and universities filter applicants through behavioral predictions made by their school AI. The virtual coach meant to support students has become their judge, their warden. And nobody talks about how it all startedwith a project that just wanted students to feel heard.

Would you like to live in one of this scenario's? Why? Why not?
I wouldnt really mind. Every technology can be seen from both a utopian and dystopian perspective. It depends on how people use it and interpret it. So I dont care too much about all thisthere will always be risks and benefits.

What happens if the technology (which you have thought of as ethically well-considered) is bought or taken over by another party?
Honestly, I havent thought much about that yet. Like many others, Im mainly focused on developing the technology in a responsible way for now. But if someone offered me a million, Id probably take itIm just being honest. I wouldnt really care what happens to it afterward if I got the money. If a business takes it over, theyd most likely change the way it works to make money for themselves. Thats how most companies operate. But at that point, it wouldnt be my responsibility anymore.

Impact Improvement: Now that you have thought hard about the future impact of the technology, what improvements would you like to make? List them below.
I guess its good to be aware of the possible future impact, but honestly, the world is constantly changing anyway. There will always be both utopian and dystopian sides to technology. Maybe we can build in some guidelines or restrictions early on, but if someone else takes over and changes everything to make money, that's on them. I can only control so much. Improvements could include clearer warnings, user limits, or opt-out optionsbut whether those survive in the long run depends on whos in charge.