Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025 ASEE
Paper ID #48275

Evaluation of an AI-based medical application using AI-generated methods: Student experiences with a case study on ”patient preference predictors”

Prof. Bernd Steffensen, University of Applied Sciences Darmstadt/European University of Technology

Studied Administrative Sciences and Sociology at the Universities in Kiel, Bielefeld (Germany), and Lancaster (UK). Doctorate in Sociology from the University of Bielefeld. Worked from 1992-2000 with Academy for Technology Assessment in Baden-Wuerttemberg

Mrs. Bettina von Römer, University of Applied Sciences Darmstadt

Bettina von Römer studied sociology at the Universities of Marburg and Bielefeld (Germany). Since 2013 she has been a lecturer at the Darmstadt University of Applied Sciences, Department of Social Sciences. Her research topics are gender studies and the impact of information technology on society.

Evaluation of an AI-based medical application using AI-generated methods: student experiences with a case study on “patient preference predictors”

Bernd Steffensen [1,2], Bettina von Römer [1,2]

[1] Department of Social Sciences, University of Applied Sciences Darmstadt
[2] European University of Technology, European Union

1. Introduction

Engineering students are often unfamiliar with ethical issues. In their actual field of study, these contents play no or at least a very subordinate role, even though technical developments and innovations often (deeply) intervene in the social contexts of their later fields of application. There is obviously a significant difference between the academic curricula and the requirements of industry. In many cases, industry demands that graduates in STEM subjects also have knowledge that goes beyond technical functioning and economic advantage. As early as 1991, the German VDI called in its paper on technology assessment: “However, functionality and economic efficiency are not pursued for their own sake. Technical systems are produced and used to expand human freedom of action. They serve non-technical and non-economic goals.” (translated with deepl) [1: p.74].

In the general discussion, this requirement is reflected, for example, in the concept of the t-shaped engineer, whose strength is seen in the great variety of interdisciplinary skills, which, in addition to mastering foreign languages, include cultural and communicative skills. In addition, young engineers are expected to think systemically and holistically, as well as to be able to critically reflect on their own actions [2], [3]. A critical examination of the concept of the t-shaped engineer and a literature review in the context of the ASEE can be found in [4].

The aim of these approaches is to lay a foundation for a technology and product development process that takes into account the non-technical and non-economic aspects addressed by the VDI in the process of technology development. Very similar approaches can be found in the “Report of the Future Ready Engineering Ecosystem (FREE)” [5: p.3-2], which was presented by the ASEE in 2024. With reference to Schwab and Davis, they mention the following key principles: “1) Focus on systems that deliver human well-being, not just on technologies; 2) Manage technologies with diverse human decision-making and agency, instead of giving in to a determinist view of technology; 3) Employ human-centered design thinking, not passive acceptance of technology as the default; and 4) Recognize values as a core feature, instead of perceiving technology as neutral and values as interference.” It is therefore about responsible action and a conscious process of weighing up the advantages and disadvantages of a technological development, with ethical aspects playing a particularly important role in many cases.

With regard to applications of artificial intelligence (AI), ethical questions in particular often come into play. The Trolley Problem, which is repeatedly mentioned and discussed in the context of autonomous driving [6], is often mentioned in this context. Another example, where the right decision can mean the difference between life and death, is in decisions about the treatment of patients who are no longer able to make these decisions themselves. This is where the AI-supported tool, the Patients Preference Predictor (PPP), comes into the picture. This software can be used when decisions have to be made in an urgent treatment situation

and the patient has not made any advance provisions for this eventuality. In such cases, an AI-based prediction of the patient's likely will is intended to support the doctors or relatives involved in the decision. Students were to address this new technical offering in a one-day seminar and examine its ethical implications.

2. Ethical considerations in the design of technical solutions

The case study to assess the ethical aspects of an AI-supported decision on patient treatment preferences (Patients Preference Predictor – PPP) was part of a course entitled "Technology Assessment in Product and Technology Development". This course is offered as a social and cultural science elective to all students at University of Applied Sciences Darmstadt. Such courses make up 5-6% of the total curriculum of all engineering degree programs at the University of Applied Sciences Darmstadt. However, students from the disciplines of design, architecture or media can also attend these courses as part of their studies. Depending on the length of the semester, the course consists of 14 sessions, each lasting 90 minutes, which are designed as face-to-face classes with lecture and discussion sections. While 10 of these sessions take place on a weekly basis, four sessions are combined into a one-day seminar on a Saturday. At the time of the day seminar, the students have already spent several weeks dealing with important aspects of technology assessment and, in particular, with evaluation criteria. They are therefore fundamentally able to create criteria catalogs that can be used to evaluate technical solutions. In doing so, it is not expected that the students will provide a comprehensive and conclusive assessment of the PPP. However, they should be able to ask a series of questions that go beyond mere technical functioning and that have been developed in line with the key principles mentioned above, as part of the FREE report.

When societies decide to use artificial intelligence applications, their use is also associated with demands from various social groups that go beyond pure functionality and economic efficiency. Deroncele-Acosta et al [7] systematize their meta-analysis of scientific articles on the use of AI in higher education into 10 pillars, one of which is AI ethics. With a very broad focus, UNESCO also addressed this topic in 2022 [8] (further references can be found in [9]). One such ethical topic is the so-called "Patients Preference Predictor". The choice of topic was prompted by the fact that the author of this article was approached by a company that describes itself on its homepage as follows: "As a leading open innovation incubator, we bring people, industries and organizations together across all borders to positively change the world with sustainable, future-oriented innovations." The young company supports open innovation projects and regularly publishes challenges for this purpose, among other things to involve universities in innovation processes. In the summer of 2024, there was a call with several sub-projects entitled "Ethical Innovation in Health Care Technology". One of these sub-projects was related to the development of a PPP. Based on extensive data analysis, the PPP helps to identify patients' presumed treatment preferences when they are no longer able to make decisions themselves. The PPP acts as a neutral and emotionally uninvolved support system. This can be particularly helpful in cases where relatives are unable to cope or existing living wills cannot be clearly applied to the current situation. Such technological support not only relieves the burden on relatives, but also strengthens the confidence that the medical care chosen actually corresponds to the values and wishes of the patient being treated. Traditional living wills are valuable instruments of self-determination [10: p. 421]. However, they also have weaknesses: Many people write living wills at an early stage and do not update them regularly. As a result, the original specifications may no longer correspond to the patient's current wishes or

altered reality of life [11]. Negative expectations about the future also often lead to very restrictive formulations, which may later prove to be no longer accurate.

In principle, it is the task of the treating party to identify the patient's presumed will and to act accordingly [12]. The PPP is being discussed as a tool to provide guidance in cases where advance directives are unclear or uncertain. By analyzing data patterns, the PPP can calculate presumed treatment preferences. However, this raises significant ethical issues. It is an AI application that has not yet been the subject of broad (scientific) discussion in Germany (an exception is Hiekel's paper from 2024 [13]). However, healthcare proxies and living wills are under discussion because they were legally reorganized in Germany a few years ago [11], [14], [15]. The legal regulation of the matter in the German Civil Code (Federal Republic of Germany) was the starting point for the students' discussion during the one-day seminar [16].

Learning objectives

The elective course in social and cultural studies at University of Applied Sciences Darmstadt is designed to teach students critical thinking so that they understand new technologies as an interrelated bundle of opportunities and risks and can make informed decisions in their future professional lives on the basis of such a basic understanding. The program is therefore one possible approach to approximating the model of the t-shaped engineer [17], [18]. The course on "Technology Assessment" (TA) was designed and taught in such a way that students learn the basics of methods for evaluating techniques. Since there is no single set of methods in technology assessment [19], the course focuses primarily on demonstrating the diversity of the various assessment criteria and on conveying that different interest groups approach the assessment of a technology with diverse value orientations and therefore arrive at very distinct assessments of the respective opportunities and risks.

The students learn about these concepts and evaluation criteria in the weekly sessions and also apply them in short group discussions. With a short homework assignment on a current technology, they apply these evaluation criteria with references to the research literature and have to advance them with a systematic argumentation to arrive at their own positive or negative assessment. These prompts for reflection open up the opportunity to apply and test their understanding according to the principles of active learning. The one-day seminar serves to conduct a critical debate on a selected example, in which students work together in groups on structured activities.

During the one-day seminar, the lecturer's main task is to moderate and clarify any questions that arise. The content or values of the various interest groups are conveyed by avatars, which are integrated into the course of the seminar via animated videos. The use of several avatars is intended to give the individual statements or positions a face, in order to make the competing perspectives and evaluations of the PPP more vivid. The expectation is that students will be better able to consider the respective ethical considerations and arguments from the point of view of the individual stakeholders.

3. The One-day Seminar

Attendance at the one-day seminar is considered mandatory and is part of the announcement and description of the course in the official course catalog of the university. The course is organized in such a way that the one-day seminar is held on a Saturday in the sixth or seventh week of the semester. The case study is organized as an integral part of this Saturday, with a

time slot scheduled from 9:00 a.m. to 4:00 p.m. This extended time span allows for a different didactic approach and the detailed treatment of a comprehensive topic.

3.1. Part one: The Morning Session – Warm up and the PPP

To start the seminar and get some movement and attention of the group the day starts after a short welcome with a group activation of a "living statistic". This activation includes 10 questions that can be answered with simple answer categories. It is about "yes" or "no", or in this case about estimation questions. To answer the questions, students have to move around the room together to get to the position that stands for the selected answer option. For this you need some free space so that the students can spread out freely. The prepared questions are then presented and the individual positions in the room for the answer options are shown. The learners take one of these positions in the room, depending on their own assessment or answer. The teacher and learners get a shared impression of the result. The aims and advantages of this method are the physical activation, at the same time there is a mutual getting to know each other and there is the possibility of an introduction to the content of the seminar topic. This substantive introduction was an important reason for using the method. The following 8 questions and 2 tasks were asked at the start of the one-day seminar. The numbers in brackets show the distribution of answers. The total number of answers varies because two students arrived late and only answered some of the questions.

1. Have you done something for your health today? Yes (7)/No (12)
2. Is the topic of health and taking care of your health important to you?
   Yes (15)/Depends (4)/No (0)
3. What do you think? How many people between the ages of 16 and 30 say in a study from 2020 (the survey took place in October/November of that year) that they think about death a lot or a great deal. (data from: [20]; green: correct answer)

27%	(10)	37%	(7)	47% (3)		57%	(0)

4. Do you know what a living will is and what it is supposed to regulate? Yes (16)/No (5)
5. Do you know whether your parents have drawn up a living will or advance directive?
   Yes (12)/No (9)
6. Have you ever considered or discussed with anyone the possibility that you might find yourself in a situation where you would have to make treatment decisions for close relatives? Yes (9)/No (12)
7. What would you say? How many respondents over the age of 16 agree with the statement that it is important to be able to rely on close relatives when it comes to making necessary decisions in the final stages of life? (data from: [21]; green: correct answer)

54%	(0)	64%	(6)	74%	(6)	84%		(9)

8. Please line up in alphabetical order by your first name. If you have the same first name, use your last name as a second sorting criterion!
9. Do you agree with the following statement?
   "Artificial intelligence is a good technical approach that we can use to solve a variety of human problems, including applications in the health sector!" Yes (21) /No (0)
10. Please line up in order of your age!

The First Avatar

The results of this introductory exercise provide a basis on which further discussion can build. At the same time, the teacher gets a first rough impression of the students' attitudes and knowledge on the selected topic. The study of the case of the PPP begins with the statement of the first avatar. This is a lawyer who explains the legal basis of the living will and healthcare proxy with brief references to the German Constitution, the German Civil Code and criminal law. After presenting the legal side, the lawyer draws the following conclusion:

> "The living will is a strong legal instrument that protects patient autonomy and clearly regulates medical decisions even when the patient lacks the capacity to consent. Its binding nature and the legally defined framework provide security for both patients and practitioners. Nevertheless, practical problems such as imprecise wording or missing updates show that continuous development of this instrument and accompanying advice are necessary. It remains an essential part of the medical decision-making process and a role model for the protection of self-determination in healthcare."

Box 1: The student's task

Imagine...

...you are dealing with a seriously injured patient who has been in an accident. We know nothing about his treatment preferences and we cannot ask him either, as he is so affected by his medical condition that he is unable to make a conscious, informed and, above all, independent decision about his treatment. The patient did not draw up an advance directive explicitly stating certain forms of treatment as desired or not desired. Close relatives are not available.

There are basically two different treatment options O1 and O2, between which a quick decision must now be made. This results in three preference options: the patient prefers O1 or O2 and – thirdly – indifference ("I don't know what is better for me"). Although the attending physician has the medical expertise to make a decision, with regard to the patient's preferences she only has the random selection. Since she doesn't know the patient, has never spoken to him, and therefore doesn't know how he would decide, she can only guess or assume. But: a decision urgently needs to be made based on the patient's state of health.

This is where an AI-based solution could help with the decision: let's assume that the patient is male, 34 years old, unmarried and has a university degree. The software of the Patient Preference Predictor is fed with statistical data that

- on the one hand is based on statistically representative surveys of a large number of people regarding their treatment preferences in hypothetical illness and accident situations and
- on the other hand on the treatment decisions actually made by autonomous and self-determined patients in comparable situations.

When evaluating the trained data and comparing it with the patient's socio-structural case data, the AI-based system concludes that 34-year-old, unmarried men with a university degree choose O1 in 80% of cases and O2 in 18% of cases, and are indifferent to the two options in 2% of cases.

From documents that the accident victim happens to have with him, we learn further statistical data about the patient, such as income and the exact content of his university education. Equipped with this additional information, the AI comes to a modified conclusion regarding preference: 75% of cases would choose O1 and 21% would choose O2, and 4% would be indifferent to both options.

We have no advance decisions from the patient and cannot let the patient decide for himself based on his state of health....

What do you think are the arguments for and against relying on the software-based solution? And above all: what could be a solution in this case?

In close reference to: Sharadin, N. P. (2018). Patient preference predictors and the problem of naked statistical evidence. *Journal of Medical Ethics*, *44*(12), 857-862.

In addition to the video of the avatars, students received:

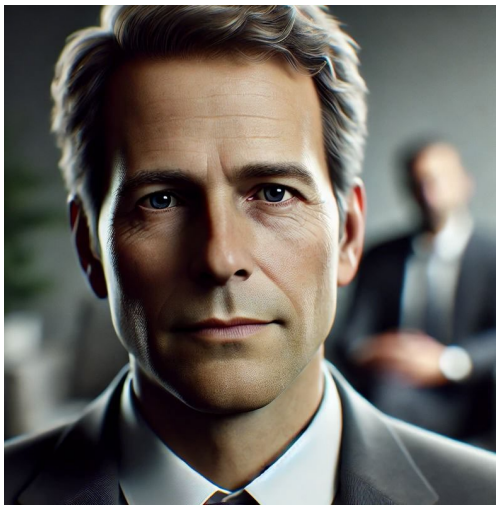
1. the statement of the lawyer in printed form. This also contains the correct information about the sources that were entered into the chatbot to generate the statement with ChatGPT. When creating the printed statement, the rules of academic integrity are meticulously adhered to,
2. the most important legal text (§1827 German Civil Code [16]) as a handout, as well as
3. a template for a living will
4. and a healthcare proxy. The two documents come from an ethically neutral well known institution that is neither ideologically nor religiously affiliated.

The students are given a little time before they are handed the actual practical case described in Box 1. Basically, the case is

taken from a text by Sharadin [22], but is slightly modified to include a few aspects to reflect the German legal situation.

The Second Avatar

Since this is the first time the PPP is mentioned during the one-day seminar, a second avatar "has its say". He is an IT entrepreneur whose company "Digital Health Ltd." has launched a PPP on the market. The company vision is generated by ChatGPT:



> "The vision of **Digital Health Ltd.** is to actively shape the future of healthcare with innovative AI-powered solutions. By developing intelligent systems such as the Patient Preference Predictor, we are setting standards for patient-centered care and individual decision support. We strive to combine technological excellence with ethical responsibility in order to sustainably relieve the burden on both relatives and medical professionals." (From the mission statement of Digital Health Ltd., generated with ChatGPT)

In addition to the company's vision statement, the spoken text contains a whole series of arguments in favor of using the PPP, as it can compensate for deficits in precautionary decision-making. The students also have a printed copy of this statement. It contains the scientific sources used to selectively filter out the advantages of a PPP. In particular, the practical experiences in hospitals and the experiences of relatives who find themselves under pressure to make decisions due to outdated or imprecisely formulated living wills show that an AI-supported assistance system like the PPP can be of help.

In order to address the aspect of technology assessment, the task given to the students was further sharpened: "What advantages and disadvantages would exist in treatment wishes or preferences identified using an algorithm? What particular challenges and hurdles could arise in practice when using the PPP to determine treatment preferences?"

The students now have some time to familiarize themselves with the documents and the arguments presented so far. When compiling all the statements that will be used during the day, care was taken to ensure that the scientific sources included are also publicly accessible to the students so that they can access the texts used directly and easily if needed.

Avatars three to five

After about an hour of discussion time, in which the working groups (each with 4 to 5 members) had gained an initial overview of the issues and the legal situation of living wills and the possibilities of the PPP, further statements were presented by three avatars. These avatars represented the following three interest groups with their video statements:

1. An emergency doctor who works in a large hospital: She reports that on many days she is confronted with having to make urgent and far-reaching treatment decisions. She briefly describes the shortcomings of existing living wills and clarifies the differences between cases of planned surgery and those in which an emergency has arisen and quick action is necessary. In cases where relatives are involved in the decision-making process and a dispute arises with the medical staff treating the patient, court

proceedings may result after the treatment. A PPP could provide helpful support to avoid such a situation.

2. The representative of the patient protection organization emphasized the importance of individual self-determination in the treatment situation. The best form of prevention remains a comprehensive and clearly formulated living will. However, the interest group recognizes the real challenges in medical emergencies when a living will is unavailable or imprecisely formulated. In such cases, the PPP can serve as a valuable tool to provide additional guidance. However, it is essential that the PPP is not used as a substitute for human judgment, but only as a supplement and support.
3. The reasoning is similar for the third avatar, who speaks for an institution that represents the interests of caregiving relatives. In addition, the question is raised as to whether it is ethically justifiable and legally legitimate to view the individual as a purely statistical case for treatment and to make treatment decisions in this way. The statement emphasizes that the use of AI systems and thus the recourse to purely statistical evidence carries the risk of neglecting emotional, moral and individual aspects.



For these statements, too, students receive handouts with the printed text. Insofar as the statements refer to statistical data (e.g. the prevalence of living wills in the population or similar), the handouts also include the complete statistics with the corresponding source references.

After a total of about two and a half to three hours, the 5 groups presented their initial preliminary results. One of the group members reported on the advantages and disadvantages of the technical approach developed by the groups, as well as the challenges. The main arguments are summarized in the following table:

Pro – advantages of PPP	Contra – disadvantages of PPP
• high probability that the personal preference will be met • second opinion for the doctor • possible legal protection for the doctor • in an emergency situation, faster decision • access to enormous amounts of data (e.g. similar cases in the past) • if no doctor is present, great help for relatives and paramedics • the more data available, the better the decision	• The AI's decision is based only on probabilities • No consideration of personal preferences • Who is liable if the AI/doctor chooses the wrong treatment option?
	• What happens if the AI and doctor's assessments differ? • The AI completely lacks personal insight into the patient • Ethical concerns that the AI makes or influences vital decisions • Is the PPP accessible everywhere – equal opportunities for everyone?

The results of the discussion that point out the disadvantages, which are shown in gray in the table, are particularly interesting. The aspect of reducing the individual human being as a bundle of statistical socio-demographic data seems unacceptable. Treatment preferences are ultimately determined by the PPP in a similar way to the purchase recommendations of the online retailer Amazon: “Customers who bought the product you are currently considering, also bought ...” In the scientific discussion, as well as in two of the five student groups, the question arises as to whether there are ways to raise the AI to a different, more ethically acceptable level: Is it possible to identify the preferences of the person to be treated more accurately than by the rather crude comparison and combination of a few socio-demographic parameters?

3.2. Part two: The Afternoon-Session – the 4P

Some scientists in the field are also calling for this next level of individualization of patient decision-making and are bringing the so-called 4P, or “Personalized Patient Preference Predictor”, into play [23]. This is intended to strengthen the autonomy of the person concerned and increase the accuracy of the treatment decision. Five options for building a P4 are presented in Earl's text:

1. **P4 trained with patients' own texts**, such as emails or social media posts, supplemented with data on their past decisions (e.g., medical treatments).
2. **P4 based on explicit responses about treatment preferences**: The model is trained with responses to questions that are recorded, for example, during systematic interviews as part of health checkups [24].
3. **P4 trained from surveys** in which individuals are prompted to reveal their fundamental values and preferences through specially designed experiments (e.g., scenarios involving different treatment options).
4. **P4 based on proxy preferences**: The model is trained with responses from relatives or proxies that predict the patient's behavior and preferences after the loss of decision-making ability.
5. **P4 trained on population-based data**: The model is trained on broad, population-based data (e.g., surveys or health data) to capture broader patterns and preferences at the population level. [25]

In order to inform students about this more advanced approach to a patient preference predictor, a press release was drafted for the company Digital Health Ltd. The company had already been introduced by the second avatar. The press release contains all the important information about the product P4 (see box 2 on the next page). In the chosen product design, the aforementioned options 1, 2 and 5 were combined. Option 1 was particularly important here, as social media activities, posts and likes are closely linked to the everyday lives of students.

As noted by the students, the use of a PPP carries the risk of reducing the individual complexity of a decision and shifting the responsibility from doctors and relatives to a technical system. In situations requiring quick decisions, this could lead to uncritical acceptance of machine recommendations, even though they may be subject to a distortion in the data or pose other problems due to the database being too limited for a preferably individualized prediction of treatment preferences.

Box 2: Excerpt from the Press Release

Digital Health Ltd., …, is launching a new, groundbreaking generation of its AI-supported system for determining patient preferences. The Personal Patients Preference Predictor (P4) is based on the latest technology and will be released as a test version in the coming days.

The previous product, the Patient Preference Predictor, has already successfully helped to determine the presumed treatment preferences of patients who are no longer able to make decisions themselves. This was based on extensive data analyses from a representative survey of 2,136 participants based on the scientific work of Rid and Wendler. The survey collected sociodemographic data as well as specific preferences in hypothetical disease situations. From this data, the system made predictions that took into account individual values and the patient's presumed will.

With the new P4, Digital Health is taking a decisive step forward: This system is based on a Large Language Model (LLM), which uses similar technologies to well-known text-generating AI systems, such as ChatGPT. The key advance lies in the P4's ability to integrate personalized data sources to get even closer to patients' actual values and preferences.

Integration of personal data for customized predictions

The P4 enables the integration of personal texts – such as e-mails, blogs or social media posts and even Facebook likes – and supplements them with additional digital information. This includes, for example, previous treatment decisions from electronic patient files, data from fitness trackers or other health-related app records. The analysis of such information helps to better understand a patient's individual values, attitudes and preferences.

Another key technical advance will be the use of modern speech recognition software. Digital Health plans to work with doctors who – with patients' consent – will record treatment sessions. These recordings will be automatically transcribed in order to integrate the information obtained into the patient's personal database.

More autonomy and self-determination for patients

The aim of P4 is to strengthen the autonomy and self-determination of patients. By analyzing the available personal data, the system filters out the values and attitudes of patients and condenses them into a decision recommendation that optimally reflects individual preferences. “With P4, we want to help align healthcare even more closely with patients' preferences and values. By combining modern AI technology and individual data sources, we are creating a new dimension of patient-centricity,” explains Dr. Julius Kempfert, CEO of Digital Health Ltd..

Future prospects

The release of the P4 test version marks the beginning of a comprehensive evaluation process to further develop the technology and integrate it into clinical practice. Digital Health Ltd. also plans to expand its collaboration with medical institutions and other research partners to optimize the application of the system in clinical practice.

The further development of the PPP towards more personalized models through so-called “fine-tuning” is intended to minimize such problems. Specific data such as medical files, surveys or even personal digital information (e.g. social media posts, fitness tracker data) could be used here [23]. However, this requires that data protection and the voluntary consent of those affected are guaranteed. However, even with an optimal data basis, the problem remains that the moral values incorporated into the algorithm do not necessarily correspond to the individual or cultural values of the patient. This also applies when LLMs are used to search personal social media data for “hidden” clues about attitudes towards life and death or treatment preferences. What do students think of this personalized approach, which uses data that younger people in particular produce daily on all kinds of topics and occasions?

Pro – advantages of 4P	Contra – disadvantages of 4P
• Personalized for each patient • Strengthening of patient self-determination, based on more personal data • Easier decision-making for doctors • Larger data volume improves preference prediction • Relatives may have more confidence in the predictions because personalized data related to the individual patient is used	• No privacy (complete screening) • Inclusion of personal data or statements that may not be included at all • How can patients decide which data should be included and used? • Who can access the data, only doctors or also health insurance companies - can the great transparency lead to disadvantages for individual insured persons (the disabled, the chronically ill)? • Are there disadvantages for people who do not produce social media data? It is not clear whether AI can answer questions about the "true will" and interests at all. It is not clear what a person's true interest is • How is the validity of the predictions checked? (Difference between decisions in an emergency and those in studies that involve decisions in hypothetical situations.)

The compilation of arguments from the student groups shows that they are indeed able to identify a number of ethically relevant criteria (comparable to [26]). It is noteworthy that questions arose in the working groups' discussions that were aimed at the inequality and disadvantage of individual population groups (older people who do not use social media, people with disabilities, etc.). In the German/European discussion context, data protection and privacy are also always of great importance. "In conclusion, the perspectives of P4 are as promising as they are concerning. Errors could have fatal consequences and bias may aggravate inequitable access to goal-concordant care. Thus, particular care will be required with regard to the design, development, implementation and continued evaluation of P4." [27: p. 36]

4. Students' Evaluation of the course

The course as a whole and in particular the one-day seminar was evaluated. This evaluation follows a standard procedure at University of Applied Sciences Darmstadt. Since only 21 students took part in the course, the statistical significance is limited. In principle, the students were very satisfied with the course and the one-day seminar and gave it an extremely positive rating. This is particularly evident in the students' overall assessment. The students' evaluation gave an average grade of 1.78 (where 1 is the best and 5 is the worst result). This result is very good compared to other courses, as the evaluation results are on average somewhat poor because many students are skeptical about this part of their elective program.

75% of students also state that learning aids of good quality are available to support learning and help to explain the facts well. This aspect was even rated with an average of 1.48. Comparable positive (1.76) is the students' assessment that the facts are evaluated from different perspectives, thus creating a well-rounded picture for evaluating the technology. The use of media, such as videos with the avatars, is rated with an average value of 1.38.

In addition to many quantitative questions, text entries are also possible for some of them. In these questions, students should give their assessments of the pros and cons of using avatars. These results are mixed. Some of the qualitative answers are given below:

- “A change from long handouts, but sometimes still not as good as videos from real experts.”
- “The AI statements were a very good idea.”
- “I liked having the video statements. For me, it's better than reading pages and pages of handouts.”
- “It can help to better and more clearly understand difficult issues.”
- “The videos helped me to better understand the different points of view.”
- “The videos were nice to have, but in the end not that relevant, because you need the handouts to go into the individual arguments.”
- “I would rather read the texts without distractions.”
- “It was sometimes difficult to follow the AI voice.”

The compilation of the various statements shows that the overall assessments are ambivalent. When considering the videos, two phases must certainly be considered. The videos seem to be well suited for getting a first impression of the topic and getting an overview. The handouts are important for more in-depth work on the topic. Especially since these can also convey more information than spoken text (additional information mentioned was statistical data or graphical representations, which can further clarify the facts in more depth).

The first statement is certainly true: the integration of real experts would definitely be preferable. However, the “Avatar experts” have the advantage that their use in the one-day seminar is flexible in terms of time. Also, creating the videos is less time-consuming than establishing contact with the experts. It is difficult to get them to make a 5-minute statement at most, and it may not be financially feasible. In particular, there are some positive aspects that point to efficiency in terms of time, finances and content. I will come back to the aspect of efficiency in the final discussion.

5. Discussion

Selwyn et al. describe how AI is penetrating many fields in education promising great potential for improving university processes (access control, automatic correction of essays or exams, tutors etc.), i.e. efficiency [28]. This assumes that unintended consequences are minimized and that developments are critically examined, deployed and used. In the present paper, the topic of AI use is addressed in two ways: First, AI-generated teaching materials (statements, avatars, videos, handouts) are used to, secondly, evaluate the AI-generated application of the Patient Preference Predictor and to analyze it in terms of its usability, social desirability, and compliance with or violation of ethical requirements. The approach described thus follows a demand by Vallis et al, who, with reference to some other authors, writes:

> “Hence, advocates of AI in education call for a stronger pedagogical and ethical approach, with more practical examples and guides for educators that are less technology-centric and more interdisciplinary.” [9: p.538]

In the critical perspective of Selwyn [28], the option is identified that we as teachers and university staff may find ourselves in a situation that Bruno Latour [29] refers to as a black box and describes as blackboxing in terms of its processuality, with small-scale processes of “educational automation”. Blackboxing is "the way scientific and technical work is made invisible by its own success. When a machine runs efficiently, when a matter of fact is settled, one need to focus only on its inputs and outputs and not on its internal complexity. Thus, paradoxically, the more science and technology succeed, the more opaque and obscure they become." [29: p. 304]. As described by Adam Smith's “invisible hand,” the way we gradually adopt

more AI tools in higher education could lead us to pursue lines of development that we did not even consider when we first started working with AI tools. UNESCO makes some critical comments in its 2023 publication on genAI in education. [30]

Latour emphasizes, in a way similar to the technology assessment in its analyses and theoretical foundations, that technology is not least the result of decisions made by both the inventors and developers and those who use a technology in the process of its creation and development. In the following, therefore, some questions about the opportunities and risks of using the described AI tools will be discussed. Since there is no well-founded scientific research to accompany the course on PPP, the insights gained are inevitably subjective, even if this subjective impression is supported by some scientific sources.

AI and software applications promise efficiency. Nevertheless, teachers and students should pay attention to the following: “1) ensuring that ethical and moral implications are addressed; 2) using AI to augment rather than replace human intelligence; 3) using AI as an instructional tool rather than a fully automated system; 4) using AI to improve academic assessment and self-assessment methods; 5) critically reviewing the results of generative AI systems.” [31] The points 1 to 3 and 5 will be addressed in the following, because: errors and biases remain that can be carried into the lecture. This raises the question of how generative AI spreads coded prejudices and perspectives and how it can be verified whether the representations of the arguments or topics are correct or superficial. Ultimately, this is a question of quality control by the instructor. Three different types of AI software products were used to create the teaching materials for the course described here:

Firstly, the two generative AIs ChatGPT and perplexity. Both have a free version, but a licensed version was used. Perplexity differs from ChatGPT in that this program provides accurate source references. The quality of the sources used can only be described as scientific to a limited extent. This also applies if the requirements in the prompt require that scientific sources are to be used by genAI when writing the text. The texts are more likely to come from general university publications that report on research activities and university news for the general public. Errors and biases can only be avoided by developing prompts that are sufficiently detailed and comprehensive. The prompts are formulated less as a text-generating task than as one with largely predefined content that merely needs to be transferred into a fluid text with a certain tenor. The efficiency of using AI lies particularly in the possibility of quickly and purposefully “beautifying” the texts in the sense of: “Write a statement for the representative of the xyz institution and use the following arguments.” In order to check the usability and content-related coherence of the text output, the teacher must acquire fundamental knowledge in advance. Without this check, hallucinations quickly take over [32]. Even though AI tools are becoming more and more powerful, human review by a teacher is still necessary to ensure text quality in cases involving such sensitive topics.

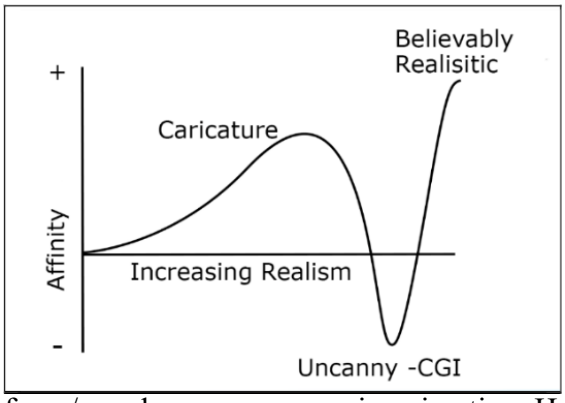
Secondly, the image generation software DALL-E: This software was used to generate the photorealistic images of the avatars. The software has an obvious bias towards “too young” and “too good-looking”. Generating images of averagely-looking “people” is comparatively difficult (this may be due to the use of German language prompts; English language prompts may work better). At the same time, DALL-E makes it possible to use the avatars harmoniously in certain contexts and to design them in a way that at least meets the lecturer's prejudices and biases.

Thirdly, the animation software “D-iD”: This software can be used to link the avatars with the generated texts so that animated spoken statements can be created that can be shown and used

as video. The software recognizes the avatar's face and moves it while it speaks, with the background remaining still. Various pronunciation errors occur in German, which is distracting. However, teachers quickly learn how to correct these (these errors tend to occur more when dealing with German texts than with English ones). One limitation imposed by the software is the length of the statement, which is around 3,800 characters. This results in videos with a length of 4 to 4.5 minutes. However, experience shows that a length of 3 to 3.5 minutes is better suited to enable students to follow the statements well.

One important final question is: How were the avatars selected and to what extent do the visual representations of these people possibly influence students' perception of the "experts' positions" and thus distort their assessment and evaluation of the respective position? This question has two components. On the one hand, it is about the quality of the graphic representation, and on the other hand, it is about the fundamental question of the perceived age of the person depicted as an avatar, their gender, their appearance or their seriousness. Are there also, for example, People of Color among the experts represented? In the courses presented here, only self-created avatars are used, which are then displayed in a suitable context/image background. Consequently, no standard options of the software programs are used. This gives more freedom in the design, but requires a little more effort. It must be admitted that in such a process, the bias of the lecturer replaces the bias of the AI. But even if real, living experts were to appear in the seminar, the lecturer would have selected them. This would also have influenced the students' opinions of the technology through the traits (likeable/unlikeable) that the experts conveyed in their statements.

Recent research by Seymour et al. [33] suggests that people prefer to interact with photorealistic, highly realistic AI-generated avatars. For a long time, this was countered by a finding called the uncanny valley [34]. Although this has not been clearly demonstrated in research [35], many studies show that the trustworthiness of an avatar for the viewer does not increase continuously along a continuum. This transition in the representation ranges from stick figures to caricatures to highly realistic representations. Rather, there is that uncanny valley, i.e. a very sharp drop in credibility when it comes to the transition to realistic representations of the avatars used. However, recent research shows that people are quite willing to trust a humanoid robot or avatar, as long as the process is transparent, comprehensible and trustworthy. [33]. This is also confirmed by the statement of a student: "If the sources are openly accessible, which was the case in the one-day seminar, a video statement is a very suitable tool."



As early as 2009, Ducheneaut et al. wrote that the increasing use and familiarization of users with digital bodies and faces has led to them becoming accustomed to avatars. "While the tools used to create these virtual personas differ widely from one world to the next, our data shows that users still tend to focus their attention on common avatar features." [36] Nevertheless, recent research shows that people are very good at recognizing errors in the representation of faces/people or even errors in animation. However, the study by Seymour et al. [37: p. 4263] comes to the following conclusion, which suggests that some of the minor errors in the animated statements that also occurred during the one-day seminar should not be seen as a reason not to use the avatars. "The results reveal that participants noticed some of the video imperfections, but this did not adversely affect their willingness to pay, affinity, or trust. We found

that once digital humans become close to realistic, users simply do not care about visual imperfections."

To what extent the representations used in the production of the videos actually go beyond the uncanny valley in terms of perceived affinity is an open question and would be a possible topic for further research.

A final question in this paper is to what extent the proposed concept can be used in other, especially larger, higher education contexts. The answer has two components.

1. The use of AI-animated avatars: These can be used in any environment in which it makes sense for didactic reasons to let an expert or an affected person (e.g. citizens of Bangladesh threatened by rising sea levels, employees of an industrial company threatened by unemployment, or students reporting stress in their field of study) express their arguments. The question is whether it would be better to use a different speaker to present their arguments or perspectives. Or is it better for the teacher to present these arguments themselves, since the deficits indicated by the discussion of the "uncanny valley" outweigh the positive benefits?
2. It also seems possible to use the case study in larger contexts, for example by using a World Café to collect arguments and opinions. This method is suitable for large numbers of participants.

In this light, minor adjustments to the operational procedure should be sufficient to discuss a case study of ethical questions regarding the application of AI in individual areas of society with students.

6. Conclusion

Initial answers were attempted to two questions: (1) what compromises have to be made when using AI-generated avatars compared to accessing real experts in the respective field (e.g. when a lecturer conducts a one-time interview with an expert and makes this video available to students)? (2) Can such avatars be used to work on an ethical issue with students? For the first question, several findings can be summarized briefly:

- The effort involved in using the avatars is significantly lower – creating input prompts compared to making contact, arranging appointments and coordinating content. This is especially true when a case study is chosen that is not in the instructor's original subject area.
- Costs for software use are significantly lower compared to the bill that might be issued if the expert were present.
- The avatar is reliable and the argumentation is clear, since this is predefined in the prompt.
- You gain flexibility in terms of time during the seminar, at least compared to the case where the expert can be questioned directly by the students during the seminar.
- The "avatar" can be reused or the statement can even be improved if, for example, the legal situation changes or new political or social discussions arise in the context of a technical development.

With regard to effects such as the uncanny valley or the influencing effect of the avatar selection, further evaluations or accompanying questionnaires are needed to show whether significant effects can be identified.

As for the second question, the use of avatars has also proven useful in the treatment and discussion of ethical issues. It must be admitted that the one-day seminar does not allow for an in-depth examination of ethical issues. However, this is also due to the fact that it is an elective subject for students who come from fields of study in which ethical and philosophical questions are not the focus of interest. The approach to the topic is rather from a practical everyday point of view.

Overall, the initial experiences with the use of avatars as experts for individual interest groups have been positive. Nevertheless, there is still room for further improvements in the design of these teaching materials.

References:

[1] VDI (The Association of German Engineers), *Technology Assessment – Concepts and Foundation. Attachment*, 2000. [Online]. Available: https://www.anstageslicht.de/fileadmin/user_upload/Geschichten/Whistleblower-Kurzportraits/VDI-RL-3780.pdf - accessed. 06.01.2025.

[2] H. Demirkan and J. Spohrer, "T-shaped innovators: Identifying the right talent to support service innovation." *Research-Technology Management*, vol. 58, no. 5, pp. 12-15, 2015.

[3] I.F. Oskam, "T-shaped engineers for interdisciplinary innovation: an attractive perspective for young people as well as a must for innovative organisations," in 37th Annual Conference–Attracting students in Engineering, Rotterdam, The Netherlands 2009, July, vol. 14, pp. 1-10.

[4] K.A. Neeley and B. Steffensen, "The T-Shaped Engineer as an Ideal in Technology Entrepreneurship: Its Origins, History, and Significance for Engineering Education". 2018 ASEE Annual Conference & Exposition, Salt Lake City, 2018.

[5] J. El-Sayed, S. DeLeeuw and R. Korte, "Preparing Engineering Students for the Future" Report of the Future-Ready Engineering Ecosystem (FREE) Workshops. Abridged Version. Washington, DC: Publication of the American Society for Engineering Research (ASEE), 2024.

[6] P. Foot, "The problem of abortion and the doctrine of double effect." In Applied Ethics, R. Chadwick and D. Schroeder, Eds. London/New York: Routledge, pp. 187-197.

[7] A. Deroncele-Acosta, O. Bellido-Valdiviezo, M. d. l. A. Sánchez-Trujillo, M. L. Palacios-Núñez, H. Rueda-Garcés and J. G. Brito-Garcías, "Ten Essential Pillars in Artificial Intelligence for University Science Education: A Scoping Review". SAGE Open, vol. 14, no. 3, https://doi.org/10.1177/21582440241272016 [Accessed Feb. 13, 2025].

[8] UNESCO, "Recommendation on the Ethics of Artificial Intelligence," Paris, 2022. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000381137. [Accessed: Jan. 09, 2025]

[9] C. Vallis, S. Wilson, D. Gozman and J. Buchanan, "Student perceptions of AI-generated avatars in teaching business ethics: We might not be impressed." *Postdigital Science and Education*, vol. 6, no. 2, pp. 537-555, 2024.

[10] T. Henking and B. V. Oorschot, "Patientenverfügung und Vorsorgevollmacht bei Krebspatienten," *Die Onkologie*, vol. *26, no.* 5, pp. 419-424, 2020.

[11] L. Stange and M. Schweda, “Gesundheitliche Vorausverfügungen und die Zeitstruktur guten Lebens, ” *Ethik Med,* vol. 34, pp. 239–255, 2022. [Online]. Available: https://doi.org/10.1007/s00481-022-00698-7 [Accessed Jan. 13, 2025].

[12] F. Nauck, M. Becker, C. King, L. Radbruch, R. Voltz and B. Jaspers, “To what extent are the wishes of a signatory reflected in their advance directive: a qualitative analysis,” *BMC medical ethics*, vol. 15, pp- 1-10, 2014.

[13] S. Hiekel, “Ein kritischer Blick auf die Idee eines Patient Preference ‚Predictors ‘”. *Zeitschrift für Ethik und Moralphilosophie*, vol. 7, pp. 333–359, 2024. [Online]. Available: https://doi.org/10.1007/s42048-024-00188-z [Accessed Feb. 14, 2025].

[14] B. Jaspers, M. Becker, C. King, L. Radbruch, R. Voltz, and F. Nauck, “Ich will nicht so sterben wie mein Vater!” (I Don’t Want to Die Like Daddy), *Zeitschrift für Palliativmedizin*, vol. 11, no.5, pp. 218-226, 2010.

[15] S. Wurm, S. M. Spuling, A. K. Reinhard and U. Ehrlich, “Verbreitung von Patientenverfügungen bei älteren Erwachsenen in Deutschland, ” *Journal of Health Monitoring*, vol. 8, no. 3, pp. 59-64, 2023.

[16] Federal Republic of Germany, “Bürgerliches Gesetzbuch – BGB §1827.” [Online]. Available: https://dejure.org/gesetze/BGB/1827.html [Accessed: Nov. 18, 2024].

[17] L. Hirst, “Transforming engineering education: creating interdisciplinary skills for complex global environments,” in Proc. of IEEE Transforming Engineering Education: Creating Interdisciplinary Skills for Complex Global Environments, pp. i-ix, Dublin 2010.

[18] C. Traver, D. Klein, B. Mikic, A. Akera, S. B. Shooter, A.W. Epstein and D. Gillette, “Fostering innovation through the integration of engineering and liberal education.” In: *2011 ASEE Annual Conference & Exposition*. 2011. S. 22.725. 1-22.725. 21. [Online]. Available: https://peer.asee.org/fostering-innovation-through-the-integration-of-engineering-and-liberal-education [Accessed Feb 14, 2025].

[19] M. Dierkes, “Was ist und wozu betreibt man Technikfolgen-Abschätzung?” WZB Discussion Paper, No. FS II 89-103, Wissenschaftszentrum Berlin für Sozialforschung (WZB), Berlin, 1989. [Online]. Available: https://www.econstor.eu/bitstream/10419/77632/1/731842421.pdf [Accessed: Jan. 14, 2025].

[20] Malteser, “Denken Sie viel über das Thema Sterben, Tod und Trauer nach?” 2020. [Online]. Available: https://de.statista.com/statistik/daten/studie/1292431/umfrage/junge-leute-zu-auseinandersetzung-mit-dem-thema-tod/ - [Accessed: Nov. 11, 2024].

[21] Berlin Institute für Bevölkerung und Entwicklung, “Was gehört für Sie zu einem würdevollen Tod bzw. zu einer würdevollen letzten Lebensphase dazu?” 2020. [Online]. Available: https://de.statista.com/statistik/daten/studie/1294673/umfrage/zustimmung-zu-bestandteilen-eines-wuerdevollen-sterbens/ [Accessed Nov. 18, 2024].

[22] N. P. Sharadin, “Patient preference predictors and the problem of naked statistical evidence,” *Journal of Medical Ethics*, vol 44, no. 12, pp. 857-862, 2018.

[23] B. D. Earp, S. Porsdam Mann, J. Allen, S. Salloch, V. Suren, K. Jongsma, ... and J. Savulescu, “A personalized patient preference predictor for substituted judgments in healthcare: Technically feasible and ethically desirable,” *The American Journal of Bioethics*, vol. 24, no. 7, pp. 13-26, 2024.

[24] A. Ferrario, S. Gloeckler and N. Biller-Andorno, “Ethics of the algorithmic prediction of goal of care preferences: From theory to practice,” *Journal of Medical Ethics, vol.* 49, no. 3, pp. 165–174, 2023. [Online]. Available: https://jme.bmj.com/content/medethics/49/3/165.full.pdf. [Accessed: Nov. 11, 2024].

[25] A. Rid and D. Wendler, “Treatment decision making for incapacitated patients: is development and use of a patient preference predictor feasible?” *Journal of Medicine and Philosophy*, vol. 39, no. 2, pp. 130-152, 2014.

[26] N. Sharadin, “Personalized patient preference predictors are neither technically feasible nor ethically desirable.” *The American Journal of Bioethics*, vol. 24, no. 7, pp. 62-65, 2024.

[27] N. Biller-Andorno, A. Ferrario and A. Biller, “The Patient Preference Predictor: A Timely Boost for Personalized Medicine,” *The American Journal of Bioethics*, vol. 24, no. 7, pp. 35-38, 2024.

[28] N. Selwyn, T. Hillman, A. Bergviken Rensfeldt, and C. Perrotta, “Digital technologies and the automation of education—key questions and concerns.” *Postdigital Science and Education*, vol. 5, pp. 15–24, 2023. [Online]. Available: https://doi.org/10.1007/s42438-021-00263-3. [Accessed: Jan. 08, 2025].

[29] B. Latour, “Pandora's hope: essays on the reality of science studies”. Cambridge Mass., Harvard University Press.

[30] UNESCO, “Guidance for generative AI in education and research,” Paris, 2023. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000386693, [Accessed: Jan. 10, 2025].

[31] T. Wu and S. H. Zhang, “Applications and Implication of Generative AI in Non-STEM Disciplines in Higher Education.” In: F. Zhao and D. Miao (eds) “AI-generated Content”. AIGC 2023. Communications in Computer and Information Science, vol 1946. Springer, Singapore. [Online]. Available: https://doi.org/10.1007/978-981-99-7587-7_29. [Accessed: Feb. 14, 2025].

[32] Y. Zhang, Y. Li, L. Cui, D. Cai, L. Liu, T. Fu, ... and S. Shi, “Siren’s Song in the AI Ocean: A Survey on Hallucination in Large Language Models.” [Online]. Available: https://ar5iv.labs.arxiv.org/html/2309.01219 [Accessed: Feb. 14, 2025].

[33] M. Seymour, L. I. Yuan, A. Dennis, and K. Riemer, “Have We Crossed the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for Realistic Digital Humans in Immersive Environments,” *Journal of the Association for Information Systems*, vol. 22, no. 3, pp. 519-617, 2021.

[34] M. Mori, K. F. MacDorman and N. Kageki, „The uncanny valley [from the field]”. *IEEE Robotics & automation magazine*, vol. 19, no. 2, pp. 98-100, 2012. [Online]. Available: https://writingstudiesandrhetoric.wordpress.com/wp-content/uploads/2023/10/the_uncanny_valley_from_the_field-1.pdf. [Accessed: Feb. 14, 2025].

[35] S. Wang, S. O. Lilienfeld and P. Rochat, „The uncanny valley: Existence and explanations.” *Review of General Psychology*, *vol. 19, no. 4, pp. 393*-407, 2015.

[36] N. Ducheneaut, M. H. Wen, N. Yee and G. Wadley, “Body and mind: a study of avatar personalization in three virtual worlds” In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1151-1160. Austin Texas USA, 2009, April.

[37] M. Seymour, L. Yuan, A. Dennis and K. Riemer, K., “Face It, Users Don't Care: Affinity and Trustworthiness of Imperfect Digital Humans.” in *Hawaii International Conference on System Sciences*, (55), 4263-4272, 2022.