

# openSAP

## AI Ethics at SAP

### Unit 1

00:00:05 Hello, and welcome to AI Ethics at SAP. My name is Sebastian Wieczorek,  
00:00:10 and I will lead you through these sessions. And we're starting with unit one,  
00:00:14 Introduction to AI at SAP. In this course,  
00:00:20 you will learn about the importance of AI ethics, how we apply it at SAP,  
00:00:25 what are the different processes, what are the things that you have to take into  
consideration.  
00:00:32 And the course will consist of five units. We're in unit one here.  
00:00:36 And we will have course assignments, and you also can discuss in our discussion forum  
00:00:42 about the contents and about the assignments. I want to start by giving you a bit of context  
here.  
00:00:51 And the context is that we, at SAP, of course, are dealing with intelligent businesses.  
00:00:56 We want to make all our customers run as intelligent businesses. And intelligence, of  
course, implies dealing with AI  
00:01:04 and implementing and embedding AI into our applications and business processes.  
00:01:11 That's what we have as a strategy at the heart of the company.  
00:01:16 And when you look at that in detail, you see that there are a lot of AI features  
00:01:22 that we're embedding into the different business processes that we have. We are looking  
across AI functions that are embedded into the main processes  
00:01:35 that are running at our customers, like lead to cash, design to operate, source to pay, and  
recruit to retire.  
00:01:43 So these are, in all of these major processes, we have a lot of AI content,  
00:01:48 a lot of AI features that are embedded. And of course, we are looking not only at how  
they're implemented in a feature correct way  
00:01:57 and in an exciting way for our customers to interact with that, but also we're taking care of  
the AI ethics aspects.  
00:02:05 And I want to give you two examples for these kinds of features. The first one that I have  
brought here  
00:02:12 is the part identification on manufacturing floor. What we see here is a very simple process.  
00:02:20 So in reality, of course, these processes might be more complex,  
00:02:25 but for the sake of explanation here, imagine that the machine part is arriving at your  
working station.  
00:02:35 As a factory worker, you're taking a photo of it because you want to know what part it is.  
00:02:40 You get an automatic classification based on this visual. And then because of that  
classification,  
00:02:50 you can detect the material number from the master code table, and you can use this  
automatically categorized part then  
00:02:59 in order to conduct the next step, like, for example, quality control, or you get information  
from it,

00:03:05 or you know what kind of manual you need to take out in order to repair something or to check something.

00:03:14 And underneath, you see what kind of business challenges we're addressing with this. I think it's quite clear that, overall,

00:03:23 we're trying to automate this machine part identification. And when you look at it from an AI ethics perspective,

00:03:35 it seems to be a very simple case. Of course, there are some ethical implications to any use of AI,

00:03:44 but when you think about it, you have a machine part

00:03:48 that is getting detected by an algorithm, and the user has to interact with it,

00:03:54 and then is finally checking what to do with the part in front of her or him. I think it's a rather uncritical process also from an AI ethics risk perspective.

00:04:10 Let's look at the second use case, intelligent candidates' resume screening.

00:04:14 Here we have a bit of a different situation from an ethics perspective. Just imagine you have a job opening

00:04:20 and you have candidates applying with a resume. What we see here is that these resumes,

00:04:26 they can come in different formats and different shapes. So what you want to do in order to automate this process

00:04:32 and to make this process efficiently, you want to digitize this information,

00:04:39 you want to extract the meaningful information. And for this, we're using two AI algorithms,

00:04:44 Business Optical Character Recognition for retrieving the digital information from the documents,

00:04:50 and then Business Entity Recognition too on this, to retrieve the important information on these documents.

00:04:58 Then, once we retrieve this information and we have figured out whether a candidate is suitable for this position or not,

00:05:05 then you can use a digital assistant to schedule a job interview and then go on with the hiring process from there on.

00:05:14 So what you see here is that the business challenge compared to the first use case is rather similar.

00:05:20 So we have digitization of information, and then we have automatic decision-making,

00:05:26 or automatic proposal of decision-making in this process. However, of course, when you look at this use case,

00:05:32 it touches on much more ethics related problems. So you have clients, you have applicants,

00:05:42 that are providing their personal-related information. You have an algorithm that is making suggestions

00:05:47 about how to deal with this information, how to retrieve that information,

00:05:51 and how to move forward in the process. And that means that from an ethics perspective,

00:05:57 there's some risk involved to, for example, discriminate or to be unfair to certain groups of candidates,

00:06:06 to not be open, and maybe limit certain degrees of freedom of candidates.

00:06:14 We have to deal with data privacy-related issues. And so from an AI ethics perspective,

00:06:20 this process poses a lot more challenges than a pure digitization process,

00:06:27 as I described it in the first case. However, what is also important to understand

00:06:31 is if we obey all these ethics guidelines, if we're careful about the process,

00:06:36 then it's a completely fine process. It's not that it has a high risk in terms of AI ethics,

00:06:43 doesn't mean that we should not implement it. It means that we have to be very careful

00:06:47 about how we implement it in a way so that from an ethical point of view,

00:06:53 this process is compliant and is acceptable for both the applicant as well as for the company.

00:06:59 And with that, I'm concluding the first session. Thanks a lot.

## Unit 2

00:00:06 Hello, and welcome to unit two, AI Ethics at SAP. What I want to talk about first is why do we need AI ethics?

00:00:14 Why are we concerned at SAP? I think you have seen the examples

00:00:19 from the previous session, and you see that there are certain aspects

00:00:22 that we want to take care of. Here in this slide you can see some highlighted points,

00:00:28 like what about human dignity in the process? What does it mean to have automated decision making in the process?

00:00:37 What does it mean to be fair to all people in the process? What kind of transparency do we need and do we want to achieve?

00:00:46 What kind of privacy issues do we need to address? What kind of privacy do we need to obtain for our employees,

00:00:54 for our people who are involved in these processes? These are all dimensions that we want to look at

00:01:00 when it comes to AI ethics. And how do we do that?

00:01:04 We at SAP created some guiding principles, and they're comprised of three different, let's say, sections.

00:01:13 In the first section, we're talking about foundational principles. You see them here.

00:01:18 The first one is saying, we are driven by our values, and what it means is that there are some base values.

00:01:26 They're not just representative for AI ethics or for AI. They're general values.

00:01:33 They are defined, for example by the Human Rights Commitment Statements. They're defined by the business ethics that SAP is applying too.

00:01:42 And we want to make that explicit that, of course, also for AI-related development, these apply.

00:01:50 Second is that we design for people. And that seems to be quite clear on first sight,

00:01:56 but it means that we're not trying to use people as an object that we want to take decisions on,

00:02:05 but that we want to put people into the center of processes and make people the driver and the benefiter of processes.

00:02:13 That's at the heart of what we are trying to achieve with putting AI into our applications and into our processes.

00:02:21 And that has a lot of implications that we then will look at also when it comes to discussing how these principles

00:02:29 then are implemented and executed. The second section is dealing with core principles.

00:02:36 You can see them here on this slide. The first one, we enable business beyond bias,

00:02:42 is dealing with the fact that, of course, we don't want to discriminate.

00:02:46 We don't want to provide AI functionalities that are supporting inequality or broadening inequality that we see.

00:02:58 We strive for transparency and integrity in all that we do. That means that we want to clearly articulate

00:03:06 how AI is used, where AI is used. We uphold quality and safety standards.

00:03:12 I think it's clear also for AI, we want to make very sure that the high quality standards that we're giving, us at SAP, are adhered to,

00:03:22 and that we also take particularly interest in safety standards that are related to the processes and tools that we're providing.

00:03:32 And last but not least, we place data protection and privacy at our core.

00:03:37 Also it's clear there, especially when you deal with AI

00:03:41 and when you deal with AI in areas where person-related data is processed,  
00:03:46 we need to pay special attention to data protection and privacy.  
00:03:51 That's clearly a core principle that we want to obey. And the third section is relational  
principles.  
00:04:00 It says here, we engage with the wider societal challenges of AI.  
00:04:04 Being experts in developing AI and integrating it into processes and applications,  
00:04:10 we, of course, are also providing expertise to the wider discussions that we have in society.

00:04:16 So we're engaging with external bodies, we're publishing principles,  
00:04:21 we're publishing our approaches, and we're consulting with a large ecosystem of  
stakeholders,  
00:04:32 be it from politics, be it NGOs, be it societal parties.  
00:04:39 Next, I want to explain how the governance is set up for AI. And at the heart of it,  
00:04:45 we have an AI ethics steering committee, and this steering committee is comprised of senior  
leaders  
00:04:50 from all around the company. So we're not just focusing on the AI functions  
00:04:55 and the teams that are dealing with these AI functions, but we're really having experts from  
all functions of the company,  
00:05:05 be it legal, be it marketing, be it development, be it quality, be it privacy, coming together to  
oversee and to develop our guiding principles  
00:05:17 and to also assess high-risk AI use cases. High risk in terms of AI ethics,  
00:05:24 as I explained already one of the cases in unit one. In addition, we set up an AI ethics  
advisory panel.  
00:05:33 This is comprised of external experts that are supporting the steering committee  
00:05:38 in definition of the guiding principles, and also helping them with advice  
00:05:46 in regard to operationalization of these guiding principles.  
00:05:51 The third pillar in our governance process is the trustworthy AI workstream.  
00:05:56 Here we have a team that established the means to implement the necessary processes in  
order to ensure compliance  
00:06:04 all around these guiding principles that are set up by the steering committee. And this is  
also comprised of people  
00:06:12 from all around the company that are volunteering to support these processes and help to  
implement them.  
00:06:21 Another important pillar of our governance approach is the AI ethics policy that we defined  
00:06:27 in the AI ethics steering committee, and rolled out via the trustworthy AI workstream.  
00:06:33 It has the purpose to clarify for employees how they relate with their daily work, how that  
relates to the overall guiding principles.  
00:06:44 So it lays out what needs to be done in certain process steps, what concrete rules need to  
be obeyed,  
00:06:52 and how the guiding principles then apply to their daily work. And this AI policy has been  
signed by the board  
00:07:01 and by the workers councils around the globe. So it now applies to every single worker at  
SAP,  
00:07:09 so everybody needs to obey this policy. But we also publish it externally  
00:07:14 so our customers and partners can see what the concrete principles are, what the concrete  
rules are that we obey  
00:07:21 and that we have given ourselves at SAP. And when you look into the policy,  
00:07:28 the main parts you can see here. They're talking about rules in regard to human agency and  
oversight.

00:07:36 So there are rules to safeguard human autonomy, especially in decision-making processes.

00:07:44 Second pillar is addressing bias and discrimination here to avoid patterns of marginalization, of inequality, and discrimination.

00:07:54 And the third pillar is transparency and explainability where we're looking at how we provide the information about AI,

00:08:05 both about the development process for AI, as well as the decisions and outcomes that are happening when we run that

00:08:14 as embedded features in our applications. Thank you.

## Unit 3

00:00:05 Welcome to unit three, where we're going to talk about the three pillars that I introduced in the last unit.

00:00:14 I want to use the use case example that I've given in unit one,

00:00:18 and without going into too many details, just for your memory,

00:00:23 we're talking here about the process where we're automatically detecting features

00:00:29 from candidate resumes in order to propose whether they're invited for a job interview or not.

00:00:38 Now, when we look into the first pillar of the AI policy, Human Agency and Oversight,

00:00:43 there are obviously more than three rules, but I want to talk about the three most important aspects

00:00:49 that the policy is covering. The first one is to enable humans

00:00:54 to override decisions of the system. Second, to choose appropriate governance mechanisms.

00:01:00 And thirdly, to avoid unintended behavior. So let's look how that translates into what has to be done

00:01:07 on this specific use case. In order to avoid unintended behavior,

00:01:12 this aspect is obviously dealing with functional correctness.

00:01:16 So it's important that the system has been tested not just for whether the intended behavior is reachable,

00:01:24 but also whether it's prevented that unintended behavior can be exposed by the algorithm.

00:01:31 So, it needs to be tested very thoroughly. When we talk about choosing

00:01:35 appropriate governance mechanisms, it's important that we have humans overseeing the process.

00:01:41 As I said in the beginning when we talked about the guiding principles,

00:01:46 we want to put the human in the center. So, it should not be the case that an AI algorithm

00:01:52 is solely taking decisions about humans. It has to be a human staying in control.

00:01:57 So a human must be able to oversee the process and the human must be able to correct the process

00:02:04 at any given time, for example, to inspect all candidate data

00:02:07 and to take decisions, and overall system decisions that are being made.

00:02:14 And that goes to the third dimension here already. It's very important that humans always have the power

00:02:23 to override what decisions have been done. Again, we want to put humans

00:02:29 in the center of our processes. So humans have to stay in control.

00:02:33 Humans have to be able to question and to override any decision that an algorithm has taken.

00:02:41 Let's look at the pillar two, Addressing Bias and Discrimination.

00:02:45 Here, we also have three main aspects that we want to look at.

00:02:49 First one, build fair and unbiased AI systems. Secondly, use inclusive data for training.

00:02:54 And thirdly, realizing measures to detect bias. So again, let's look at how this translates to the use case.

00:03:02 When we're using inclusive data for training, it means that the data set that we're using

00:03:07 has to be representative of the users and of the people that are comprised

00:03:15 and interacting with the algorithms. it means we have to make sure that the data

00:03:19 that we're using in this HR process is covering all kinds of user groups that we're expecting

00:03:28 to use this algorithm. So, we have to see that they're coming

00:03:32 from all different regions of the world, that ethnic minorities are represented,

00:03:39 that skill sets are represented in a meaningful way, that age groups are represented,

00:03:44 gender is represented, and so on and so forth, so that in the end we're creating an algorithm

00:03:49 that is not biased and is focusing solely on the qualifications of the applicants.

00:03:58 We want to build a fair and unbiased AI system. For that, it's not just important to make sure

00:04:04 that the training data is representative, but also that the training process itself

00:04:10 and the framework in which we're integrating the system are also not introducing bias.

00:04:17 So as you probably know, not only training data, but also the algorithmic part

00:04:25 may determine whether people are getting discriminated against or whether bias or aspects of fairness are not considered.

00:04:34 And for that, we also then need to take care and look at the system as a whole,

00:04:41 not just at the training data. And of course, when we've implemented all that,

00:04:46 we also need to realize measures to detect bias when the system is then in action.

00:04:51 So, we have to make sure that when the systems are running, we continuously track and monitor

00:04:58 that the system is not exhibiting behavior that is unwanted, especially in terms of fairness.

00:05:10 The third pillar is Transparency and Explainability. Again, there are three key aspects,

00:05:15 documenting data sets and the development process, documenting also the AI system's

00:05:21 capabilities and limitations, and providing transparency about AI-generated output.

00:05:27 So again, let's look into how that translates into concrete requirements for the given use case.

00:05:34 When we talk about data sets and development processes, it's very important that we're documenting

00:05:40 what data has been used, how it has been interacted with in the development process,

00:05:45 how, for example, the data formatting, the sampling, the data labeling has been taken care of,

00:05:51 and what algorithms have been applied to this data set. Then once we have trained the algorithm,

00:05:59 so we can put it into the product, of course, we need to also document

00:06:04 what are the exact capabilities of this algorithmic part and what are the limitations of it.

00:06:10 So, how well is the algorithm behaving? What kind of precision does it have?

00:06:15 What kind of recall does it have? In order to create a better understanding

00:06:21 of what you can expect from the application and what you cannot expect from the application.

00:06:28 And it has to be done in a way that non-technical people understand these limitations clearly too.

00:06:35 And thirdly, also, when the system is running then, when you're interacting with the system,

00:06:41 the AI-enabled system should provide you with information about why certain outputs have been delivered,

00:06:49 why certain suggestions are being made. Like, for example, what are the features that the candidate is exposing

00:06:56 that make this candidate interesting for going forward in the process?



00:07:02 So this kind of information, again, needs to be provided in a way that the users are able to understand it

00:07:09 and then take appropriate decisions so that we have a human-led process.

00:07:15 And with that, I'm concluding the discussion about this unit.

00:07:20 Thank you.

## Unit 4

00:00:05 Welcome to this unit where we're talking about the operationalization of the AI ethics policy

00:00:11 that we introduced in the previous unit. The first thing that I want to talk about

00:00:16 is the AI Factory process that you see here on the left-hand side of the slide.

00:00:21 It starts with ideation. So in the first phase, people are coming together

00:00:26 to discuss what's the use case, what's the business value,

00:00:29 what could the product feature look like. So basically, ideation of the AI component,

00:00:38 as well as how it is embedded in the product. The second phase is the validation

00:00:42 where, on the one hand, data scientists are looking into data

00:00:46 that they can use in order to realize the AI function, but on the other side,

00:00:52 also on the business side, people are interacting with the customer, for example,

00:00:56 to see what would an integration look like, how would this feature be feasible

00:01:01 and desirable and viable? Third phase is the realization phase

00:01:07 where, basically, after all these experiments are conducted, all the feasibility studies have

00:01:14 been conducted,

00:01:14 the AI features are developed and are productized so that they can be integrated then in the

00:01:24 product,

00:01:24 which is happening in phase four, the productization. So developers that are concerned

00:01:31 with the business logic of the products are taking then these AI features and components,

00:01:37 and integrating these into the business process they're responsible for.

00:01:41 And phase five then is operations, where people are maintaining, supporting,

00:01:48 running these services, if there are services in the product,

00:01:53 so that customers can utilize the AI features and the products that are surrounding them.

00:02:00 And obviously, this process is not like in a waterfall model,

00:02:05 where you go through one step after the other, but it's an iterative process

00:02:09 that can go back to ideation at any point in time. Or it can go to the previous steps

00:02:15 if there needs to be refinement. And I'm showing you this AI Factory process

00:02:22 because it's a good way to explain what kind of considerations that are defined in the AI

00:02:31 policy

00:02:31 have to be taken into consideration at which step because obviously, in ideation phase,

00:02:36 you have to look at different ethics-related questions than in the productization or in the

00:02:45 operations phase.

00:02:45 Next what I want to do is I want to talk about the different stakeholders

00:02:50 that are involved in this process. You already heard from the definition of the process

00:02:55 that a lot of different people are coming together in the different phases in order to work

00:03:01 together.

00:03:01 Obviously, you have an AI use case owner, so someone who's responsible for the project

00:03:07 overall

00:03:07 and is coordinating the activities, and this person is going to lead

00:03:13 through the whole process. But then you have other roles

00:03:16 that I'm going to introduce now briefly that are coming in in certain phases

00:03:22 and are not active in other phases or just adjacent in other phases.

00:03:27 Like, for example, the AI data scientist, whose task is to do the validation

00:03:34 and maybe also sometimes they're involved in the productization of what they have been building

00:03:39 in terms of AI features. The AI engineer,

00:03:43 which is more active in the realization phase, then after the experiments are conducted

00:03:50 in order to build out the actual AI feature. Product developers that are responsible

00:03:59 for integrating these AI features then into the products. Or here on the second side,

00:04:05 designers that are concerned with the user experience, so how does the user interact

00:04:11 with the features. User assistance experts who are providing

00:04:21 things like documentation, for example, or maintaining training information.

00:04:29 We have AI OPS engineers, which are then responsible for operating the AI functions

00:04:36 that are provided. And of course, we have the customer,

00:04:39 who is all these steps, hopefully interacting with the development teams

00:04:46 and with the other people as a stakeholder, representing the user of the AI solution.

00:04:54 And why did I show you this? First of all, I wanted to tell you how broad the policy is

00:05:01 in terms of in which steps they're applicable, and policy basically concerns all the different steps

00:05:08 of the development lifecycle. And it's also relevant for all these different personas

00:05:16 that I have been introducing. Of course, it can be possible that in smaller projects,

00:05:23 people are wearing multiple hats so that, for example, the designer is also responsible

00:05:28 for user assistance, or the AI data scientist is also responsible

00:05:34 for the tasks that are related to AI engineering. But in general, all these roles exist

00:05:43 even though they're mapped to the same people. And for all these roles, there are specific actions

00:05:50 that need to be taken care of, and considerations that need to be taken care of

00:05:56 in the different phases. And in order to help people to understand

00:06:02 what I have to do in a certain role, in a certain process, we created the AI Ethics Handbook

00:06:08 in which you can find for each of the phases that I explained, and for each the personas

00:06:15 that I explained, what are the responsibilities, and what are the things that need to be taken into consideration

00:06:23 in order to comply with the policy. And then by that, also comply to the guiding principles of AI

00:06:31 that SAP has been setting up. Thank you.

## Unit 5

00:00:06 Hello, in this unit, we will be talking about how to assess the risk of AI use cases,  
00:00:11 in terms of AI ethics. Here, you can see an overview of categories  
00:00:17 of AI ethics-related use cases, or AI ethics use cases, so to say.  
00:00:22 There are cases that are red-line use cases, which means that they should not be  
conducted at all.  
00:00:28 There are cases that pose high risk in terms of AI ethics and here we will talk about how to  
identify such use cases.  
00:00:38 And then there are standard use cases that don't pose a high risk, that are handled in a  
standardized way, then.  
00:00:46 And we will also look at what this standard procedure is for use cases. And when you think  
about high-risk use cases,  
00:00:55 in unit one we introduced the HR use case, you can go back to that unit to listen to that.  
00:01:02 We also have an example for a standard use case, which was the parts detection use case  
  
00:01:08 also presented in unit one. So let's talk about red-line use cases first.  
00:01:16 Obviously, SAP has certain standards in terms of ethics that we want to obey and that we  
don't want to cross.  
00:01:25 And that means that use cases that touch certain red lines should never be followed up at  
SAP.  
00:01:36 For example, SAP does not want to work on human surveillance, does not want to engage  
in any use cases that are set up for discrimination,  
00:01:49 or does not want to follow up on use cases that have the aim to de-anonymize people.  
00:01:57 That's when you look at personal-freedom-related red lines. When you look at society-  
related red lines,  
00:02:04 we don't want to engage in manipulation of public opinion. We don't want to undermine  
public debate,  
00:02:14 and we don't want to intentionally harm users, or harm societies in providing systems that  
are harmful.  
00:02:27 And on the third line, we also of course, don't want to damage the environment, so we don't  
want to engage in AI systems and build AI features  
00:02:39 that can provide explicit danger to the environment. This is not...when you look at it, I think  
it's clear what the direction is,  
00:02:50 and for me, it's also important to say that these are examples for what red lines are. There  
are more red lines that you can define  
00:02:58 based on the code of conduct that SAP has signed, we also discussed that in a previous  
unit.  
00:03:05 Please look at also other business-ethics-related guidelines of SAP. So if you have a use  
case that is touching any of these red lines,  
00:03:18 basically this use case has to be stopped, also in accordance with the AI ethics policy.  
00:03:25 When we look at high-risk use case, as said, we made this HR use case example there,  
00:03:32 the obvious question is, how do I figure out whether a certain use case that I'm involved  
with, that I have in front of me,  
00:03:40 is a high-risk use case or not? And for that, we created five simple questions that have to be  
answered.  
00:03:47 If you answer yes to the first question, and to any of the questions two to five,  
00:03:54 then the use case is considered a high-risk use case, which means that the steering  
committee needs to engage  
00:04:03 in some kind of evaluation and approve the use case. So let's look at the questions.

00:04:09 Question one, which is the most decisive question, if it's answered no, you're in a standard use case,

00:04:18 is asking you, are you processing, or is in the use case any personal data processed?

00:04:25 And that means both during the development time, so is person-related data used for training an algorithm,

00:04:33 as well as during design time, so is person-related data put into the system

00:04:39 in order to compute a certain outcome? If the answer is no, it's a standard use case.

00:04:47 If the answer is yes, then please answer the questions two to five.

00:04:53 And if you have a second yes, then as I said, the steering committee needs to go into an assessment.

00:04:59 So question two is the question for protected personal data, or special categories of data, as it's called in many legal obligations.

00:05:12 So we're talking about sexual orientation information, information about religion, biometric data.

00:05:18 So that may include face images, for example, that can be used for face recognition.

00:05:25 Is that processed? If yes, then have the steering committee assessment.

00:05:31 Third question, could the outcome of what the algorithm is providing negatively affect humans? So is there a chance that fundamental rights, like personal freedom, are affected?

00:05:46 If yes, then the steering committee needs to approve the case.

00:05:52 If there is automated decision making in place, which means that the algorithm or the system surrounding it

00:06:00 are taking automated decisions moving forward a process, then the steering committee needs to approve this case.

00:06:09 And fifth question, if the use case is embedded in a high-risk sector, like healthcare,

00:06:15 like law enforcement, like HR, then again, the steering committee approval is needed

00:06:23 in order to move this use case forward. And what does the process look like

00:06:29 if you need approval from steering? You can see here on this slide.

00:06:35 So the check of or the answering of the questions is obviously the first thing that a use case owner has to conduct.

00:06:43 For this, we created a form that you can fill out and get this information. And then the first step is that the use case owner,

00:06:50 after determining yes, it's a high-risk use case I'm engaged in, is creating a use case description

00:06:57 and sending this to the AI ethics office for further processing. Then the AI ethics office in the next step will set up a due diligence.

00:07:07 So experts from the AI ethics field are then engaging with the use case owner, figuring out to what extent the policy has been obeyed,

00:07:18 answering together all the questions that are necessary. And then in the third step,

00:07:25 the use case owner will provide these findings in a short summary to the steering committee, answer in a steering committee session to the questions

00:07:35 that the steering committee members have. And then the steering committee in the end and the last step

00:07:41 is providing binding guidance. And binding guidance could mean that the use case can be continued as presented

00:07:51 or that the steering committee is asking for improvements on the use case.

00:07:57 For example, on transparency aspects or on anti-discrimination aspects.

00:08:04 Or it could be the case that the steering committee is deciding that the use case needs to be stopped

00:08:10 because the risk is not acceptable. For everything that is not a high-risk use case,

00:08:18 so a standard use case, the use case owner can go through the due diligence on her or his own.

00:08:28 So again, the first step, the mandatory step is to check the risk.

00:08:32 But if the risk level is low, so yes has not been provided as an answer two or more times, then the use case owner is checking for compliance,

00:08:44 documenting the use case and the risk assessment, and then providing that information on request so that it can be reviewed.

00:08:54 That's then all that is necessary, so no involvement from the AI ethics office or the steering committee needed.

00:09:03 And with this, I'm concluding the session overall. What you have learned is how important it is to apply AI ethics

00:09:14 and how important it is for SAP that people are aware of it and are following the guiding principles and the policy.

00:09:22 You learned how the policy is set up, how it is mapped to the development and deployment process.

00:09:34 You have learned how a risk assessment is conducted and how the process of review is set up at SAP.

00:09:45 I hope you had fun and learned something new. And with that, thank you, and have a good day.

## Unit 6

Language subtitles  
(translated by machines)

00:00:05 Hello. In this unit, we're going to talk about the ethics of generative AI.

00:00:12 First of all, I would like to talk about what foundation models are, which are the basis of generative AI.

00:00:17 They come in different flavors, so they can process text, image, speech, structured data, and all types of other data.

00:00:26 The specialty or the interesting part about foundation models is that they can be used to provide multiple capabilities with the same model.

00:00:36 For example, large language models that are used to provide text, they can be used for summarization, for content creation, for translation,

00:00:46 search, code generation, and all kinds of text related activities. And they're working out of the box, meaning that you don't need to tailor them

00:00:57 to specific tasks or create copies and train them on these specific tasks, but they can out of the box provide these capabilities.

00:01:05 Also, they're very easily accessible for end users because they're interacting in natural language.

00:01:12 And if you want to know more, please also consider taking the OpenSAP course on AI and Generative AI at SAP.

00:01:23 Generative AI is associated with a range of risks. On the one hand traditional AI ethics related risks,

00:01:32 but then also specific risks that are related to GenAI and to foundation models.

00:01:38 And we will look at these GenAI specific risks here in this unit. So when we look at it, these foundation models

00:01:47 come with a large number of technical limitations, even though they are very cool in providing all these

00:01:55 additional capabilities. So for example, when you think about math and counting, these models are not very

00:02:02 sufficient in providing you good answers. You all know that these models have issues with hallucinating, plausible sounding,

00:02:11 but false information. There is a lack of determining the accuracy or the confidence

00:02:18 of the output and it's also challenging for providers of large language models and other foundation models to keep them up to date.

00:02:26 So they need to be constantly retrained, given that also information about the world and context is evolving.

00:02:34 And when you look at the examples here, you can see that we asked who are the founders of SAP in a model, I think it's GPT 3.5.

00:02:45 And you see here that this is an example of hallucination where three times Hasso Plattner has been called out, which obviously

00:02:53 shows that fact-based interaction is a weak spot for foundation models in general.

00:03:01 There are also non-technical risks associated with GenAI. For example, users who look at these plausible sounding answers take them

00:03:09 for granted without thinking about the hallucination risk. People misusing these models to generate fake content is a huge issue.

00:03:21 Detecting bias in unstructured data resulting from the unsupervised training, So meaning that you have uncured datasets that nobody has been looking

00:03:35 into is an issue. And of course, there are also legal issues related to IP and copyright.

00:03:42 And here on the bottom, you see what kind of copyright issues may arise. In this example, there has been a foundation model

00:03:52 that disclosed training data directly based on prompt. Now here you can see on this slide  
some of the specific considerations for GenAI

00:04:06 and foundation models. When you look into transparency and explainability,  
00:04:11 for example, it's clear that it's hard to provide sufficient information about the training data  
and the architecture

00:04:19 because most of the providers of large language models keep this information secret and  
it's often proprietary information that is not

00:04:28 disclosed to the public. When it comes to bias and discrimination, one of the issues  
00:04:34 is that because foundation models are creating unstructured data, it's much harder to detect  
bias and discrimination in this output

00:04:45 compared to structured data like tabular data where you can use statistical methods to  
detect these.

00:04:52 And when it comes to human agency and oversight, it's clear that generative AI, and  
specifically large language models,

00:05:00 can easily provide concise and relevant and well-sounding, understandable explanations for  
a certain text

00:05:10 or for a certain output, but of course it's not actually providing the information about the  
process of generation,

00:05:18 because that's inherent to the model and even the model is not able to understand those.

00:05:26 So in conclusion, SAP combines the power of generative AI with business data and  
business processes.

00:05:34 So we're at SAP using GenAI in our applications in combination with the information that we  
get from our systems.

00:05:44 We keep the human in the loop and make the human responsible for the output and also for  
generative AI.

00:05:51 We are following the same principles of AI ethics as we are following for all the other AI-  
related processes.



© 2023 SAP SE or an SAP affiliate company. All rights reserved.  
See Legal Notice on [www.sap.com/legal-notice](https://www.sap.com/legal-notice) for use terms,  
disclaimers, disclosures, or restrictions related to SAP Materials  
for general audiences.

