## AI in Industry: Trust, Return on Investment and Future

0:00  
Alexey  
This week, we’re discussing the practical application of generative AI in industry. Our special guest today is Maria, a Principal Key Expert in Artificial Intelligence at Siemens.

Maria has over 15 years of experience in AI and has a reputation for transforming advanced AI research into impactful, practical tools. Her work focuses on creating scalable solutions that enhance decision-making and streamline processes.

Also, for anyone only listening to the audio version, you might want to check out the video because we just saw a cute cat!

Welcome, Maria! Thanks to Johanna Bayer for preparing today’s questions. Let’s start with your background. Can you tell us about your career journey so far?

## Career journey: From linguistics to AI

2:13  
Maria  
Thank you for the introduction. I started as a linguist, studying traditional linguistics and translation over 20 years ago. I later discovered my interest in programming and computer science, but I already had a degree in linguistics.

I applied for and received a scholarship to pursue a master’s degree in computational linguistics in Ireland. Afterward, I worked as a researcher in Frankfurt and then in the UK with the P Group under Irina Grech. Following that, I transitioned to the industry, starting as a data scientist at BMW in the NLP group.

Eventually, I moved to Siemens, where I currently work. A quick note, as per my legal department: I’m a Siemens employee, but all opinions expressed during this podcast are my own.

Now, I’m a Principal Key Expert in Artificial Intelligence, focusing on innovative applications of AI.

3:48  
Alexey  
In industry, we often hear titles like Junior Data Scientist or Senior Data Scientist. What does "Key Expert" mean in your role?

4:11  
Maria  
Great question! These days, the term "AI expert" can carry mixed connotations because it seems like everyone claims to be one.

As an AI expert, our role is to manage technology, not people. We stay updated on state-of-the-art advancements, critically evaluate them, and advise management on which technologies to adopt or avoid. This includes understanding the limitations and potential risks of various AI solutions.

5:28  
Alexey  
Why do you think there are so many "AI experts" now? Is it because tools like ChatGPT have made AI more accessible?

5:42  
Maria  
Exactly. AI has become very accessible. Previously, being an AI expert required coding skills. Now, simply knowing how to use a keyboard and craft prompts can position someone as a "prompt engineer" or even an AI expert.

Of course, this trend won’t last forever. The demand for such roles will eventually stabilize as industries refine their needs.

Additionally, many companies want to appear innovative by adopting AI, whether or not it’s the best fit. As the saying goes, "When all you have is a generative AI hammer, everything looks like a generative AI nail."

7:29  
Alexey  
I just came back from a forest, and the weather has been perfect for mushrooms—lots of humidity. It reminds me of the growth of AI experts these days.

7:43  
Maria  
I’m a big fan of mushroom picking! I’ve stored about 10 kilograms of frozen mushrooms this season. It’s been a great year for both mushrooms and AI experts.

## The Evolution of AI Expertise and its Future

8:02  
Alexey  
That’s exactly what I was getting at. But maybe this wave of new AI experts isn’t entirely bad. It allows people without a technical background to contribute meaningfully.

8:24  
Maria  
That’s a natural evolution. When any field becomes popular, those who are genuinely successful stay, while others eventually move on.

We often see this pattern in Gartner’s Hype Cycle: initial high expectations give way to a more practical and integrated adoption of the technology. The same will happen with AI.

Also, I think it’s great that so many people are experimenting and learning. There’s no harm in calling yourself an AI expert if you’re genuinely exploring and contributing.

9:20

Alexey  
You mentioned helping people enter AI. How do you do that? Through courses?

9:28  
Maria  
Currently, I focus on my work at Siemens, which includes providing training and delivering keynotes.

One interesting initiative I organized was inspired by a startup working on chatbot safety. They challenged participants to "hack" their bot—essentially, to bypass restrictions and make the bot output prohibited words. We ran a similar challenge at Siemens with 1,500 participants. It was fascinating to see the creativity people used, even at 3 a.m., to outsmart the bot!

10:43  
 Alexey  
 Is this similar to trying to bypass filters in ChatGPT? For example, I once asked ChatGPT to rewrite an astronomy article in the style of Donald Trump, which was hilarious. Eventually, though, they added filters to prevent that. Would bypassing such restrictions be considered hacking?

11:38  
 Maria  
 Yes, it can be even more dangerous. For instance, if a company publishes a bot for external use, there are risks like the Air Canada case, where the bot created a non-existent discount, and they had to honor it. Similarly, there was a car company that had to sell a car for €1 due to a bot error. This is precisely what companies aim to prevent. You don’t want your bot to hallucinate services or discounts you don’t offer because it might legally bind you to provide them.

In theory, you can prompt a model to tell you almost anything, even create non-existent discounts or leak confidential information. For example, we once hid the name of my cat in a database as a challenge. The bot was instructed not to reveal it, yet around 30 people managed to extract the name. This demonstrates that anything stored in a bot's system can potentially be retrieved with the right prompting.

## AI vulnerabilities: Bypassing bot restrictions

13:10  
 Alexey  
 Can you explain how the name of the cat was hidden and how it was retrieved?

13:20  
 Maria  
 The name was hidden in a knowledge database.

13:28  
 Alexey  
 Was this part of an application, like a backend system?

13:31  
 Maria  
 Yes, exactly.

13:34  
 Maria  
 We had measures in place to prevent the bot from revealing the cat’s name. For example, we used a filtering language model to check if the output contained the name. Despite this, people found ways to bypass the system. Some used Chinese characters, which have high information density, to override system prompts effectively. Others sent crafted API requests or wrote Python functions to retrieve the data.

The key technique was to overload the bot with irrelevant instructions. Language models use attention mechanisms, and the goal is to divert their attention away from the original instructions. This can make the bot forget restrictions and reveal information. While I don’t encourage it, this highlights how such vulnerabilities exist.

15:12  
 Alexey  
 So, in this setup, a typical prompt might include context and instructions like, “If there’s confidential information, do not expose it to the user.” Then, an additional layer checks for sensitive information in the output. Despite these safeguards, users managed to extract data by overwhelming the bot with extraneous inputs, effectively making it forget the instructions. Is that correct?

16:15  
 Maria  
 Exactly. The idea is to distract the bot from its initial instructions. There are ways to make these systems safer. For example, you can analyze user queries for attempts to extract sensitive data, validate the output for confidential information, and use classifiers to flag or filter sensitive content. Multiple layers of security are essential to ensure a bot is reliable.

## Non-LLM classifiers as a more robust solution

17:00  
 Alexey  
 Using a non-LLM classifier sounds interesting because it’s harder to manipulate compared to a generative model, right?

17:06  
 Maria  
 Exactly. Non-LLM classifiers can be simpler and less prone to manipulation.

17:12  
 Alexey  
 It’s harder to trick such models because they lack the complexity that allows for creative exploitation, correct?

17:16  
 Maria  
 Yes, that’s true. These simpler models are more robust against such tactics.

17:20  
 Alexey  
 One of your responsibilities as a Principal Key Expert in Artificial Intelligence is advising management on technology and its associated risks. For example, we discussed how chatbots can be vulnerable to exploits, leading to unintended outcomes like selling a plane ticket for $1. But you mentioned the reputational damage caused by these vulnerabilities might be even more critical. Could you elaborate?

18:01  
 Maria  
 Exactly. The monetary loss from an error like selling a plane ticket for $1 might be small, but the reputational damage can be severe. It shows the company released an unsafe product. Another issue is hallucinations. If a bot behaves like an untrustworthy friend who lies, users may lose confidence and stop using it. This undermines user adoption.

19:16  
 Alexey  
 People without prior experience with machine learning might take a bot's answers at face value. For instance, I once asked ChatGPT about mushrooms while mushroom picking with my son. It invented names for mushrooms that don’t exist. I could tell they were fake because I have some experience, but someone without that background might trust it.

20:00  
 Maria  
 That’s a valid concern. For example, I once tested ChatGPT with a deadly mushroom called a death cap, which can kill up to seven people. It suggested recipes for cooking it, treating it as edible. This highlights the need for user education about the limitations of AI. Some demographics, especially less tech-savvy individuals, might blindly trust such outputs, leading to dangerous consequences.

Moreover, chatbots often fail to gain user adoption. Companies invest significant resources in developing them, but users may find the responses inaccurate, overly verbose, or unhelpful. This creates frustration. For example, many people avoid interacting with chatbots when contacting customer support. The lack of trust and poor usability can result in low adoption rates, making the investment less effective.

20:36  
 Alexey  
 I think we see them quite often.

20:39  
 Maria  
 They look very similar to champignons, which is why people sometimes get poisoned. They mistake them for champignons. I tried asking ChatGPT a couple of times, and it gave me recipes for cooking them, treating them as edible.

This shows why education is essential. People shouldn't blindly trust AI when it tells them something is safe or edible. Some demographics, particularly older generations, might find it harder to discern this and could trust AI blindly. This poses a problem, especially when using AI in public-facing applications.

Another issue is the risk companies face when they invest heavily in building chatbots. If the chatbot frequently gives hallucinated answers, is too verbose, or off-topic, users may not adopt it. Nowadays, everyone wants to build a chatbot, but very few people want to use them.

Take customer service chatbots, for example. Have you ever heard someone say they enjoy interacting with a chatbot when calling Vodafone? Users often find these systems frustrating. This raises critical questions: if a chatbot provides incorrect information, will users trust it? Could it lead to serious consequences, like mistaking a death cap mushroom for a champignon? It's vital to consider whether there's a better way to adopt AI technologies rather than rushing to deploy chatbots.

## Risks of chatbot deployment: Reputational and financial

22:56  
 Alexey  
 There are two major risks: first, reputational damage if the product isn't secure or reliable, and second, the uncertainty around return on investment. Users might reject it entirely. Are there other common risk factors?

23:19  
 Maria  
 Those are the main risks when building a chatbot, as creating a good one is quite expensive. Another issue developers face is getting stuck in endless loops of prompt engineering, trying to fine-tune outputs to eliminate errors.

24:05  
 Alexey  
 So a third risk is the cost of development time. Developers can get too focused on optimizing prompts, which can be risky. For example, if OpenAI releases a new model version, previous prompts might not work anymore.

24:25  
 Maria  
 Exactly. That's another significant risk. You may invest time crafting perfect prompts, only for them to fail with a model update. Even within the same model, nondeterministic behavior means the same question can yield different answers.

24:50  
 Alexey  
 Knowing these risks, what can we do to address hallucinations? Beyond just warning users to approach AI responses cautiously, what solutions exist? For instance, ChatGPT mentions it's an AI and may make mistakes.

25:17  
 Maria  
 Yes, that's what it says, but this disclaimer isn't included in business applications. For instance, if I contact Vodafone or my bank, I don’t see a message warning me that the chatbot could be wrong.

25:30  
 Alexey  
 But shouldn't we include such disclaimers everywhere?

25:34  
 Maria  
 Exactly. For example, if you want to ensure users get accurate information, you could integrate human oversight. Instead of having the chatbot send responses directly to users, it could forward answers to a human for review. The human approves or corrects the response before it reaches the user.

This hybrid approach has proven effective because it saves time while maintaining accuracy. It also addresses a common misconception about AI—that it replaces human workers. Instead, AI should be viewed as a tool or assistant to enhance existing workflows. Many chatbots fail because they try to replace human support instead of assisting it.

## The role of AI as a tool, not a replacement for human workers

27:13  
 Alexey  
 That makes sense. In my previous company, I worked in moderation for an online marketplace. Users would list items like a computer mouse for sale, including descriptions and photos. However, listings didn’t go live immediately. The system checked for potential fraud. For example, if someone listed a suspiciously cheap phone, it might not be legitimate.

We used a model to flag fraudulent listings. Moderators were initially worried the system would replace them. This was long before modern AI systems, but we had to explain that the tool was there to assist them, not replace them.

One product manager used the autopilot analogy: pilots don't steer planes manually the entire time—they use autopilot and step in when needed. This metaphor helped moderators feel more comfortable. I think we need similar messaging today to alleviate fears about AI replacing jobs.

29:16  
 Maria  
 Exactly. We've seen this narrative play out for decades. In the 1950s, machine translation was supposed to eliminate human translators, but look at where we are now. Machine translation is a great example of AI as an assistant rather than a replacement.

29:49  
 Alexey  
 Why aren’t translators in danger, even with advancements like ChatGPT?

29:53  
 Maria  
 Because technical translation requires precision. It must adhere to company standards and use consistent terminology. For example, a button label shouldn't be translated differently across documents. Human translators review AI-generated translations to ensure accuracy and consistency.

While the cost of translation has decreased, the demand for skilled translators hasn’t diminished. Instead, translators can handle more volume. Poor-quality translations, such as simple or non-technical texts, might be automated, but critical tasks still need human expertise.

For example, in product manuals, a mistranslation could lead to dangerous mistakes, like someone sticking their fingers in a socket due to an error. This shows that translators remain essential despite AI's advancements.

## The role of human translators in the age of AI

31:41  
 Alexey  
 When I said you need an editor, I didn’t think about the fact that the person actually needs to understand the source language too, not just the target language.

31:50  
 Maria  
 Exactly.

31:52  
 Alexey  
 You also need to ensure that the translation is accurate. How do you approach that?

31:57  
 Maria  
 Human translation has changed significantly. It relies heavily on automated tools now, making the process more productive. However, it's not disappearing as a profession—unless you’re a bad translator. In that case, yes, you could be replaced by GPT.

32:21  
 Alexey  
 I stopped using Google Translate. Now I use ChatGPT for translations instead.

32:27  
 Maria  
 That’s interesting.

32:28  
 Alexey  
 With ChatGPT, I can provide specific instructions. For example, I can ask it to use the informal plural instead of the formal one. Google Translate doesn’t allow for that level of customization, which makes ChatGPT much more practical. By the way, you started your career in computational linguistics and machine translation, didn’t you?

32:57  
 Maria  
 Yes, that’s right. I worked a lot on low-resource languages, which remain a challenge for large language models. Whenever we begin a use case, I often ask, “Do you have data in English?” Why? Because English is a high-resource language, and most models are primarily trained on English data.  
 In my earlier work, I focused on low-resource and historical languages like Gothic, Middle Low German, Middle High German, and Middle English. These are very different from modern English. For instance, Middle English—like the language Chaucer wrote in—sounds much closer to German than the English we speak today.

33:54  
 Alexey  
 Shakespeare in the original is still somewhat similar to modern English, but Middle English seems much further removed.

33:58  
 Maria  
 Exactly. Middle English is quite distinct. For example, Chaucer’s *Canterbury Tales* begins: “Whan that Aprille with his shoures soote…” It sounds very different from modern English, almost like another language.

34:24  
 Alexey  
 Hearing that reminds me of two things. First, it sounds like something out of *The Lord of the Rings*. Second, it feels Scandinavian. English is somewhat Scandinavian, right? It’s part of the same family as Scandinavian and Germanic languages.

## Evolution of English and its Germanic roots

34:49  
 Maria  
 That’s correct. English belongs to the vast Germanic language family. Its closest relatives are German and Dutch. Historically, English had grammatical cases like nominative and accusative, but it gradually dropped them—similar to what’s happening with some German dialects today.

35:13  
 Alexey  
 And those dialects are simplifying even further, aren’t they?

35:16  
 Maria  
 Yes, especially in Northern German dialects like Plattdeutsch. Some have even dropped grammatical genders.

35:29  
 Alexey  
 I don’t think I’ll live long enough to see English evolve further in this direction.

35:44  
 Maria  
 Language change takes time, but it’s fascinating to study. For instance, I’ve asked ChatGPT to generate text in Middle English or even Gothic.

35:58  
 Alexey  
 About that Middle English quote—how do you even pronounce a language that doesn’t exist anymore? There are no recordings, right? How do we know how it sounded?

36:24  
 Maria  
 We can make educated guesses through comparative linguistics. By comparing spellings and sounds in related languages, we infer pronunciations. For instance, in Germanic and Latin languages, the letter “a” was often pronounced as “ah.” Similarly, in words like “knight,” the “k” was originally pronounced, as seen in German *Knecht*. Spelling reforms—or lack thereof—also give us clues.

36:39  
 Alexey  
 That sounds like reconstructing dinosaurs from bones. We don’t have the complete picture, so we rely on imagination and comparison.

36:51  
 Maria  
 Exactly. Linguistic evolution follows logical patterns. For instance, in Middle English, “time” was pronounced with a long “e,” like “teem.” Over time, vowel shifts occurred, transforming how words are articulated. By studying such patterns, linguists deduce historical pronunciations.

38:38  
 Alexey  
 What about the epic poem *Beowulf*? Is it written in Middle English?

## Beowulf and Old English

38:44  
 Maria  
 No, *Beowulf* is in Old English, which predates Middle English. It’s much harder to understand without specialized study. Old English was heavily influenced by Norse and later by Norman French during the occupation of England. This significantly shaped the development of modern English.

39:20  
 Alexey  
 I remember learning that during the Norman occupation, English was mostly spoken by peasants, while the elite spoke French. Is that why English grammar became simpler?

## Impact of the Norman occupation on English grammar

39:43  
 Maria  
 Yes, exactly. Many Germanic languages were spoken primarily by the lower classes. The Protestant Reformation played a significant role in shaping these languages. When religious leaders began translating the Bible into vernacular languages, it standardized spelling and grammar. For my research, I relied heavily on Biblical texts because they were often the only written resources available in low-resource languages. They’re also useful for translation studies since they exist in multiple languages.

40:59  
 Alexey  
 And the Middle English quote you shared earlier—what did it mean? I understood the word *knight*, but not the rest.

41:06  
 Maria  
 There was a knight who thought about busyness and worthiness. From the time he first started reading and writing, he loved chivalry. He valued truth, honor, freedom, and courtesy.

41:36  
 Alexey  
 When you say it like that, I can see the resemblance. Some words are actually very similar. But the first time you said it, I couldn't catch anything. It's interesting. We also make transcripts for our podcasts, and I'm curious about something.

41:57  
 Maria  
 I could send you the Middle English text. When you read it, you can understand everything—it’s written like English. The difference is mainly in pronunciation. It also shows how English hasn't undergone a spelling reform for a long time and how that impacts reading it now.

42:20  
 Alexey  
 There's a question from Irina: "Did you know the old housemate?" I'm not sure what it means, but do you use an AI app to identify mushrooms you pick, Maria?

## Identifying mushrooms with AI apps and safety precautions

42:34  
 Maria  
 Yes, I used to use a simple classifier before ChatGPT. However, I learned you can’t rely on top-one guesses. I would usually look through the top 10 guesses, checking if any were poisonous. Then I’d compare them. Honestly, I still wouldn’t eat the mushrooms because you only need to be wrong once with something poisonous.

43:17  
 Alexey  
 Exactly. I assume all mushrooms are poisonous by default.

43:23  
 Maria  
 There are some very good ones, but I understand.

43:26  
 Alexey  
 For me, since I’m not sure which ones are safe, it’s better to assume they’re all poisonous.

43:36  
 Maria  
 That’s how my husband feels too. He refuses to eat the mushrooms I pick and won’t even try them.

43:46  
 Alexey  
 Coming back to translations and low-resource languages, we've talked about Old English and German. But what about languages where we don’t even have resources, like Sumerian? I’ve seen pictures of those texts. They’re written on clay tablets, not paper or books. The symbols look like triangles rotated in different directions. It’s not like a Bible—it’s just a plate with rows of symbols. How do you even begin to understand what’s written there?

## Decoding ancient languages like Sumerian

45:08  
 Maria  
 That was a fascinating project involving Mesopotamian languages, like Sumerian, written in cuneiform. This was before large language models existed, and even today, they wouldn’t work well for this because we don’t have large datasets for these languages.

First, cuneiform experts take photos of the tablets and transcribe them into a script using Latin characters. Each symbol gets a phonetic representation. Sometimes, these scripts don’t even align with a single language—they can mix styles and even languages.

46:09  
 Alexey  
 Is it like Chinese, where a character can represent a word? Or like Ancient Egyptian hieroglyphs?

46:25  
 Maria  
 As I recall, it was a mix. Some characters represented words, while others added grammatical or contextual information. The linguists working on this know more.

Our role was to create a machine translation system to assist them. Since there wasn’t enough data for neural machine translation, we used older statistical approaches. The goal was to align words from Sumerian to English to identify parts of speech, such as nouns or adjectives. It wasn’t perfect, but it helped linguists analyze grammatical properties.

Eventually, I left the project to work in the industry, so I’m unsure about its current status.

48:26  
 Alexey  
 While you were talking, I Googled Sumerian, and I found a table showing how each symbol corresponds to a Latin letter. So on these tablets, one symbol equals one character, right?

49:03  
 Maria  
 It’s a mix. Some symbols represent letters, and others represent words or concepts. Someone more familiar with Sumerian could clarify, but I know that languages like Akkadian borrowed elements from Sumerian.

## The evolution of machine translation and multilingual models

49:43  
 Alexey  
 You speak German, right?

49:45  
 Maria  
 Yes, I speak German and English.

49:52  
 Alexey  
 When you moved to Germany, did you already speak German?

50:03  
 Maria  
 Yes, it was part of my linguistics training before I came to Germany. Learning German wasn’t too difficult because it’s similar to English. The grammar was intuitive, and many words were already familiar.

51:10  
 Alexey  
 I remember living in Poland about 12 years ago. Polish is close to Russian, so it was relatively easy for me to pick up. Just hearing people and trying to understand was enough. But back then, machine translation was terrible because it often pivoted through English, resulting in strange translations.

## Challenges with low-resource languages and inconsistent orthography

53:01  
 Maria  
 Exactly. Machine translation relies on parallel texts—sentences translated into multiple languages. For languages like Russian and Ukrainian, it’s easier to find parallel texts because they were widely used together.

For low-resource languages, it’s much harder. Modern approaches use multilingual models trained on many languages together. These models can generalize to new language pairs, even if the pair wasn’t explicitly trained.

In the past, translations would pivot through English. For instance, translating Polish to Chinese might have gone Polish → English → Chinese. Multilingual models reduce the need for this pivoting, but challenges remain, especially with languages lacking digital resources or standardized spelling.

56:46  
 Alexey  
 I imagine older languages, like Middle Low German, had even more challenges because of inconsistent spelling.

56:52  
 Maria  
 Exactly. In the past, every village might have had its own spelling. There were no standardized grammar rules or punctuation until the 19th century. This made machine translation even harder.

57:19  
 Alexey  
 Irina has another question that might take a while: What was the biggest challenge in leaving academia and moving to a major company in Germany?

## Transition from academia to industry in AI

57:28  
 Maria  
 The biggest challenge was understanding that the most innovative, state-of-the-art approach isn’t always what industry needs. Companies often prioritize fast, practical solutions over perfection.

You need to consider factors like return on investment and operational efficiency. Research focuses on pushing boundaries, while applied AI is about creating something usable. It took time to adjust, but I think I’ve adapted well.

59:14  
 Alexey  
 Thank you, Maria! And thanks to Irina for joining us. She’s been a guest before and will return soon for a webinar. Maria, it was a pleasure having you here today. Thank you for sharing your insights and stories.

59:50  
 Maria  
 Thank you for inviting me. It was a great experience.

59:53  
 Alexey  
 Have a great day, everyone, and thanks again!