



## Delivering socially acceptable AI: Beware the subtleties

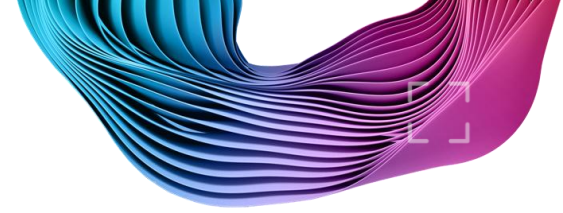
To successfully scale AI, executives need to ensure the technology doesn't break societies' written and unwritten rules. Jan-Jan Lowijs outlines how organisations can ensure their AI systems are accepted by customers, employees, regulators and policymakers.

Like human beings, artificial intelligence (AI) must comply with laws, regulations and other rules. Activities that are against the law, morality or public order without AI, are also against the law, morality or public order with AI.

But compliance isn't as easy as it sounds. Firstly, AI can break rules faster than humans do. Secondly, humans are very bad at making rules, so we often need to divert from them. But more importantly, much of what we consider 'normal behaviour' is not laid down in written rules of any kind.

Yet, if an enterprise AI system behaves badly the company's reputation and brand will suffer. Conversely, AI systems that are too timid can be inefficient, which could curb a company's competitiveness. Consider this theoretical example. You are riding your bike at 4am and you encounter a red traffic light. There is no other traffic in sight, no cameras and no observers. Most people wouldn't stop, but an AI system would follow the letter of the law.

This dynamic is playing out in the real world. For example, the developers of self-driving cars are wrestling with the question of how assertive their AI systems should be. During a recent power outage in San Francisco, the traffic lights at various junctions failed. Whereas human drivers proceeded carefully, self-driving Waymo vehicles didn't cross those junctions, thereby obstructing traffic. After facing criticism for sticking rigidly to traffic laws, Waymo vehicles are now under fire for driving in the style of an "aggressive New York taxi driver". Getting this balance right is clearly proving tricky.



## The importance of etiquette and empathy

Another challenge for AI is understanding how humans are nice to each other, respecting etiquette and common courtesy, while showing empathy. When someone is clearly upset, other human beings seek to reassure and comfort them. To be accepted, AI systems also need emotional intelligence.

Good manners and respect are hardly ever captured in law or even policy, yet most people adhere to those rules because they learnt how to behave around other people as part of their upbringing and education. Lacking the benefit of 18 years of parenting, AI systems are prone to errors of judgement, such as generating photos of celebrities in their underwear.

Ironically, immature AI systems can also be too empathetic. Take the example of a North American airline offering discounted tickets to people travelling to funerals. When a prospective passenger asked whether he could claim the discount after travelling, the airline's customer service AI agent responded positively. But when the customer applied for the reduction, the airline said the request needed to be submitted before the flight. Challenged in court, the airline lost. The court held that whatever the AI agent says, it says on behalf of the airline and is, therefore, official.

## Teach AI systems how to behave

How can organisations ensure that their AI systems respect both the codified rules and the uncoded rules that govern our lives? Even answering the first part of this question isn't straightforward. AI systems need to comply with AI-specific rules, such as the EU's forthcoming AI Act, and broader legislation governing human beings and their equipment.

The first step is to identify exactly which rules are applicable to the task AI is being asked to perform – if your AI sorts resumes, it needs to respect employment laws, but it doesn't need to take traffic law into consideration. However, this scoping exercise isn't necessarily a one-off – AI systems are designed to evolve over time, so be aware of mission and scope creep.

Note that laws vary by country – what may be compliant behaviour in one country, can be a violation of a law in another. For example, free speech in one jurisdiction might constitute libel in another. Location matters!

Ensuring AI systems respect the unwritten rules and cultural norms in each geography is harder still. The first step is to ask yourself whether you would be comfortable explaining the AI system's behaviour on prime time television news in the jurisdiction where it operates. If not, then that indicates the system is contravening human norms. Another test is to ask how you would feel if the AI system was interacting with a member of your family. There are other similar questions that can help you establish whether an AI system is socially acceptable.

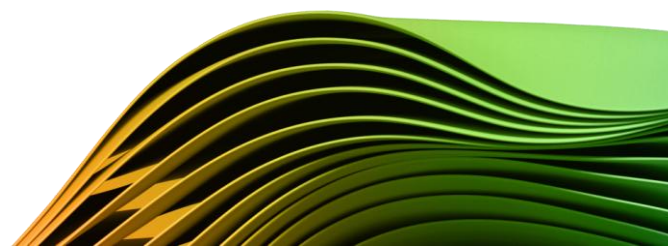
In general, it is important to be transparent about why an organisational AI system is behaving the way it is and then enable customers or citizens to opt out of the relevant interaction. If 99% of people opt out, then you know the system is contravening social norms. If 50% opt out, then you need to call on the judgement of other humans.

To help achieve the right balance between using AI to achieve an immediate business objective and protecting the organisation's reputation, you can subject the system to review by a group of your peers or the organisation's ethics board.

## Prepare a contingency plan

Even when you are confident that your AI system is respecting both the written and unwritten rules of the jurisdictions where it operates, don't get complacent. Assume that something will go wrong and prepare a contingency plan that will allow you to take the AI system down and give your organisation the time required to fix any glitches.

It is also essential to revisit your assumptions regularly. Social norms evolve as new generations become adults and technology moves on. What may have been unacceptable ten years ago, may be acceptable today and vice versa. Don't assume an AI system will detect these subtle changes – always apply human judgement.



In summary, it is not enough to simply ensure that your AI systems don't break any laws. To safeguard your organisation's reputation, it is vital that these systems treat employees, customers, suppliers and other stakeholders with respect and sensitivity. Ethics aren't optional. Like humans, AI systems should live by a tried and test life rule – treat others in the way that you want to be treated.

## Contacts

To find out more about how to manage AI-related risk, contact us.



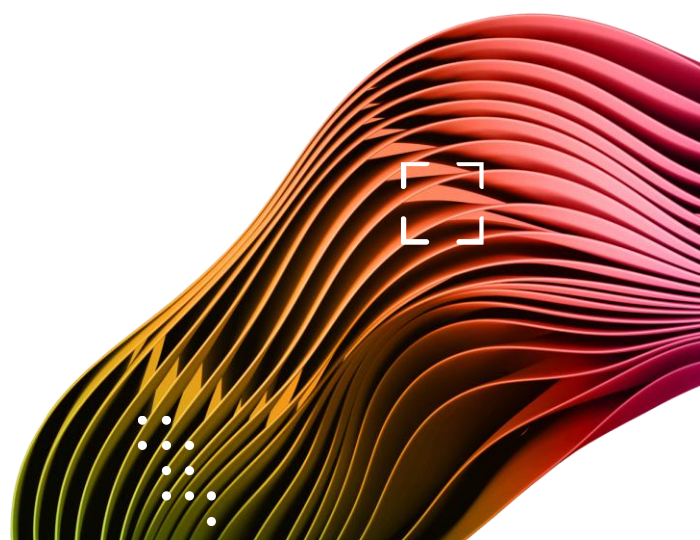
**Jorg Schalekamp**

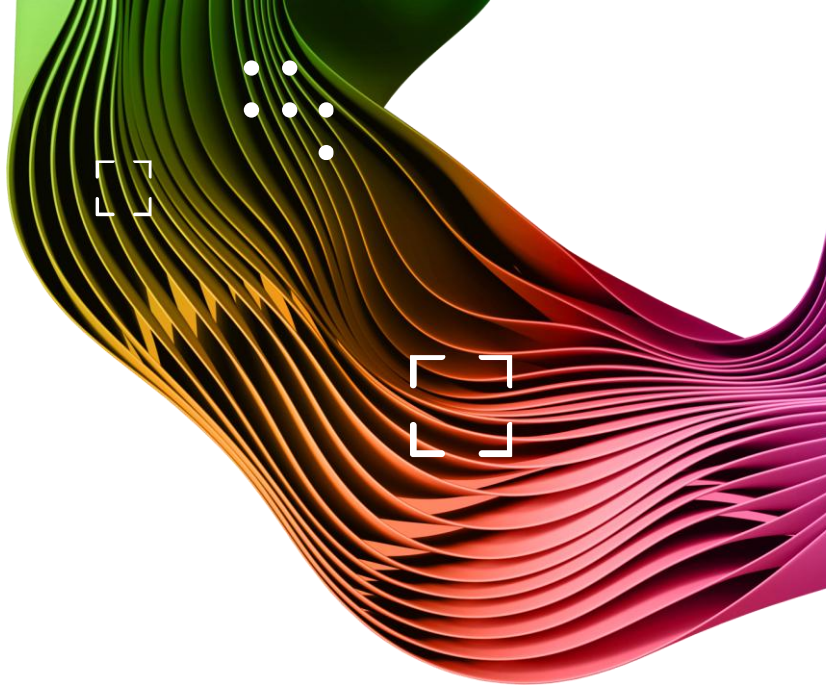
Partner Engineering, AI and Data & AI Leader  
[JSchalekamp@deloitte.nl](mailto:JSchalekamp@deloitte.nl)



**Jan-Jan Lowijs**

Director Privacy & Digital Regulations Services  
[JLowijs@deloitte.nl](mailto:JLowijs@deloitte.nl)





# Deloitte.

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited (“DTTL”), its global network of member firms, and their related entities (collectively, the “Deloitte organization”). DTTL (also referred to as “Deloitte Global”) and each of its member firms and related entities are legally separate and independent entities, which cannot obligate or bind each other in respect of third parties. DTTL and each DTTL member firm and related entity is liable only for its own acts and omissions, and not those of each other. DTTL does not provide services to clients. Please see [www.deloitte.com/about](http://www.deloitte.com/about) to learn more.

For Netherlands use: ©2026 For information, contact Deloitte Netherlands.

CoRe Creative Services. RITM2402012