# Creepy Trolley Problem: Algorithms, Inequalities, and Safety as Privilege

**Mehmet A. ORHAN, Ph.D.**
**EM Normandie Business School**
**Paris Campus, France**

**Sep 1, 2026**

**CASE STUDY**

**Keywords**

AI ethics, algorithmic control, dystopia, face recognition, inequalities

# Creepy Trolley Problem:
# Algorithms, Inequalities, and Safety as Privilege

## INTRODUCTION

Every day, our service experiences are shaped by interactions that involve subtle hierarchies. These hierarchies may stem from disparities in opportunities and privileges. In restaurants, regulars could receive special treatments, such as personalized attention, complimentary extras, or guaranteed tables without reservation, even during peak hours. Likewise, familiar faces at nightclubs might bypass the queue with little more than a nod from the bouncer. While some of these privileges could be relational, they could also be transactional. For example, don't want to wait in long lines at airport security checks? Or want to board before anyone else for your flight? Purchase an upgraded ticket with fast-track options to skip unnecessary waits. Many of these small luxuries are available for a price. However, what we know is that these advantages are rarely neutral because people are constantly assessed, either *consciously* or *unconsciously*, and treated differently based on their fame, familiarity, social or economic status, perceived importance, or potential (future) value as customers.

Even when these transactions are operationalized, human involvement still matters. Yet, more and more services are being delegated to machines. One example could be that robotic waiters are already replacing human service providers in some restaurants. They might identify returning guests, but could they independently decide to offer something extra? And if so, how might they be empowered to take such initiative? Under what conditions should they be allowed to offer preferential treatment at all? Or, how about the common experience of being slightly over the baggage weight limit in planes? A human agent might show mercy and waive the fee, recognizing the stress of travel; but would a machine, which is programmed for strict compliance, do the same? Without being explicitly coded to allow for such 'human' exceptions, the algorithm remains indifferent to the context of the situation.

Machines are now involved in many parts of the service industry. In the age of artificial intelligence and autonomous systems, humans increasingly rely on new evolving technologies to navigate daily life. Algorithms do not yet "know" us in the human sense, yet as technology advances, they are learning to assess our behaviors, preferences, and even our social standing. In fact, this process may already be underway. Some of these developments have been very revealing…

Maybe, you may have noticed how generative algorithms respond differently to identical prompts depending on your previous interactions with them. Some algorithms have already started evaluating and treating us differently, just as humans do. The question, then, is how we can build trust and meaningful relationships with systems that observe, categorize, and make decisions about us. What happens when these systems are empowered to make choices with moral or ethical consequences, decisions that could directly affect our safety, well-being, or even our lives? And, more importantly, could safety, a fundamental human right, be commodified in the same way as extra legroom or priority boarding? And, what if we aren't even customers? How does our relationship with algorithms shift when we're merely bystanders, outside any formal business relationship, yet still subject to their judgments, inferences, and decisions?

Some of these questions may sound philosophical, but more and more, they are becoming part of our daily concerns. This case study aims to bring one of the most fundamental philosophical questions, the Trolley Problem, into a genuine practical one with a touch of dystopian perspective. Considering how algorithms, robots and self-driving autonomous vehicles have been advancing in the last decade, critics warn that the commercial incentives driving their deployment may carry profound consequences for social equality and democratic accountability.

This case study examines the ethical dilemma faced by Alex Hughes, a senior software architect at DriveSafe Systems, one of the world's fastest-growing self-driving vehicle manufacturers. With over a decade of experience in safety-critical automotive software, Alex has built his reputation on designing systems that are both technically robust and ethically sound. He serves as head of the Algorithm Integrity division, giving him direct authority over modifications to DriveSafe's core navigation and decision-making codebase. Outside work, Alex is a father of two and a committed cycling advocate in his community, someone who thinks often about road safety not as an abstract principle, but as a lived reality. When a team of licensed third-party mechanics, certified DriveSafe partners operating in an emerging market, approach Alex with a proposal to alter the company's self-driving algorithms specifically for wealthy, famous, and influential individuals, including celebrities, athletes, and politicians, the request lands squarely within his area of authority. Although the likelihood of accidents involving these high-profile individuals is extraordinarily slim, the mechanics assure their clients the utmost safety, whether they are pedestrians, drivers, or passengers. Moreover, the team believes that this strategic move could provide them with a competitive edge in the fiercely competitive car servicing industry around the world. The team also speculates that customers could

be interested in such technologies as they increasingly sought personalized and exclusive services in every aspect of their lives. What began as a technical adjustment thus carried profound ethical weight, raising questions not only about market advantage but also about fairness, equality, and the responsibilities of those who design life-and-death technologies.

## BACKGROUND

Across industries, rapid technological transformations are reshaping our daily experiences, and transportation is no exception. Self-driving cars have introduced a new dimension to the age-old moral dilemma known as the trolley problem. This problem is a classic thought experiment in moral philosophy that explores ethical decision-making in life-and-death situations. It presents a scenario in which a runaway trolley is hurtling down a track toward a group of people who cannot move out of the way. We, the observers, have the ability to intervene by pulling a lever to divert the trolley onto another track. The trolley problem has thus become relevant in modern debates about autonomous systems, such as self-driving cars, where machines may need to make split-second decisions that could result in harm. In these cases, the ethical questions are no longer purely abstract and theoretical, considering how much we rely on algorithms these days…

Self-driving cars operate through sophisticated camera systems and sensors that continuously scan the environment, detecting objects, humans, other vehicles, and traffic situations in real time. These technologies allow the vehicle to navigate safely and respond instantly to hazards that would require split-second decisions from a human driver. However, as face recognition and social evaluation technologies advance, these systems could do more than simply detect obstacles. They could begin to assess the social value of the individuals in their path, raising profound ethical dilemmas about who lives and who is endangered in critical situations.

The origins of this case study are not completely artificial, and the ethical questions surrounding autonomous systems were not entirely theoretical. The rise of surveillance technologies in some of the world's most technologically advanced societies bears the paradox that the governments that have long been considered the most democratic are today increasingly deploying systems that monitor citizens at unprecedented levels, and these practices often contradict basic principles of privacy and human rights. Thus, this case study aims to assess complex issues governing today's technological landscape with a critical lens. It thus aims to bring the discussion of social hierarchies into autonomous vehicles, resonating with these wider trends, highlighting the tension between technological progress, ethical responsibility, and the erosion of fundamental rights.

**The Original Trolley Problem**

The trolley problem is a classic thought experiment in moral philosophy. A runaway trolley is heading down a track toward five people who cannot move away. The observer has the option to pull a lever, diverting the trolley onto a side track where one person is tied down. The dilemma asks whether it is morally acceptable to sacrifice one life to save many. Variations of the problem explore whether active intervention is ethically different from inaction and how factors such as age, occupation, or social role influence moral judgments.

What's more interesting today is that self-driving cars are often equipped with multiple cameras embedded inside, allowing 360-degree visibility inside and outside the vehicle and making many deliberate decisions to provide a smooth transportation experience. With advanced technologies such as complex algorithms, these vehicles are also capable of making instant decisions that can have certain ethical implications. And, many of these implications have been part of ongoing debates. On the other hand, despite the availability of cutting-edge technologies like facial recognition, some features remain underutilized, opening avenues for further ethical considerations.

**The Creepy Trolley Problem**

One autumn afternoon, Alex was approached by a team of mechanics operating in the Gulf region, specifically the United Arab Emirates. They had an unusual proposal. The mechanics suggested modifying the company's self-driving algorithms so that vehicles would prioritize the safety of high-profile individuals, i.e., wealthy clients, celebrities, athletes, and politicians. The rationale was straightforward: such customers would pay premium fees for assurance that, in the unlikely event of an accident, the car's algorithms would make every effort to protect them first, whether they were pedestrians, passengers, or drivers.

The mechanics framed this as a competitive opportunity. They believed that in the fiercely contested UAE automotive services market, one of the most lucrative in the world, driven by ultra-high net-worth individuals, government clients, and an appetite for cutting-edge technology, offering elite algorithmic protection could provide a decisive competitive edge. The UAE had already positioned itself as a global testing ground for autonomous vehicles, with pilot programs running in Dubai and Abu Dhabi, and a regulatory environment that, while evolving, remained more permissive than those in Europe or North America. The prospect of gaining influential clients and lucrative contracts tempted them, and they speculated that customers might welcome such privileges if the technology remained discreet.

**The Challenge**

The premium service the mechanics envisioned was not limited to passengers inside the vehicle. This was its most unusual dimension. Under their proposal, a subscribing client's biometric profile, that is, their facial geometry, would be registered in a secure cloud database linked in real time to every enrolled self-driving vehicle. As those vehicles moved through traffic, their external cameras, already continuously scanning the surrounding environment, would simultaneously run facial recognition checks on every person in their field of vision: pedestrians on the pavement, cyclists at junctions, occupants of other cars, bystanders at the roadside. The moment a registered face was detected, the vehicle's decision algorithm would reprioritize its collision-avoidance logic to protect that individual above all others, regardless of physical proximity or conventional rules of the road. In effect, a wealthy client walking home from dinner or standing at a bus stop could, without any interaction with the vehicle, trigger a protective override in a car they had never entered. Safety, in this model, would travel with the client's face.

Crucially, the mechanics did not present themselves merely as clients seeking a feature request. They were certified DriveSafe service partners with deep access to the vehicle systems they maintained, and they had already demonstrated, internally among themselves, that the underlying algorithm could be modified through targeted exploitation of the software update pathway they used for routine maintenance. In plain terms, they had the technical capability to introduce the changes unilaterally, bypassing official development channels entirely. What they lacked was the insider knowledge to do so cleanly, without triggering the integrity checks and audit logs embedded in DriveSafe's core codebase. That was precisely why they had come to Alex. As head of Algorithm Integrity, he knew exactly where those safeguards lived and how they could be circumvented. They were not simply asking for his approval. They were inviting him into a collaboration, offering him a share of a lucrative and discreet business, while making it quietly clear that his cooperation would determine whether the modification was done well or done badly. The choice, they implied, was his.

Beyond the UAE, the mechanics had a larger ambition. They asked Alex a pointed question: could this service scale globally? They were keen to understand whether the facial recognition and priority-protection model could be extended across different markets and jurisdictions; Europe, North America, Southeast Asia, wherever self-driving vehicles were gaining a foothold. More specifically, they wanted Alex to explore the regulatory, technical, and commercial feasibility of a worldwide rollout: which legal environments would be most permissive, which markets harbored the highest concentration of ultra-high-net-worth individuals willing to pay a premium, and crucially, whether DriveSafe's

existing global software infrastructure could support the real-time biometric database synchronization that a multinational deployment would require. They were not asking Alex to commit to a global operation overnight. They were asking him to map the terrain, to quietly assess the appetite and the obstacles, while the UAE pilot remained their proof of concept. For Alex, this additional request transformed the nature of the proposition entirely. What had begun as a localized, if deeply troubling, business arrangement was now being framed as the foundation of a global commercial ecosystem built on algorithmic privilege.

Modern autonomous vehicles that are currently circulating in traffic integrate multiple technologies to ensure safe navigation. Sensors and cameras provide 360-degree visibility, detecting objects, vehicles, pedestrians, and road conditions. At the same time, radar and lidar measure distance, speed, and the shape of surrounding objects. Computer vision and machine learning algorithms process visual and sensor data in real time to classify hazards and make navigation decisions. Decision algorithms then select actions such as braking, swerving, or continuing straight, often within fractions of a second. Theoretically, current technological advances also allow additional features such as facial recognition and behavioral assessment, raising the possibility of assessing the identity, social value, or perceived importance of individuals in critical situations.

For Alex, the proposition raised some concerning questions. Modern self-driving cars already employ cameras, sensors, and advanced algorithms to scan environments and make instant decisions. Adding facial recognition or social evaluation could expand capabilities further. But it also risked embedding social hierarchies into life-and-death decisions. Who should be spared in an unavoidable accident? Should wealth or fame factor into the code? What should be the price for it?

But these decisions are not straightforward. Should technology mirror existing social hierarchies, or resist them in the name of equality and human dignity? Many critics have already highlighted a range of ethical challenges associated with autonomous driving technologies: First of all, biases and inequalities in society are a major concern. Programming that values certain individuals over others risks embedding social privilege into safety decisions. Moreover, while national and international laws may prohibit discriminatory algorithms and raise serious liability issues, regulatory enforcement remains uneven, especially in regions with weaker oversight. Authoritarian governments, paradoxically, are among the most aggressive adopters of surveillance and social-scoring technologies, even as democratic societies increasingly deploy mass monitoring systems that challenge privacy and human rights.

**The Ethical Dilemma**

Alex now stood at the center of an ethical dilemma. On one hand, implementing the mechanics' proposal promised significant business advantages, prestige, and market differentiation. It could also strengthen partnerships in an expanding global region. On the other hand, it directly conflicted with principles of fairness, equality, and public trust.

If the public learned that self-driving cars treated lives unequally, outrage and regulatory backlash would be inevitable. Governments might introduce stricter oversight, lawsuits could emerge, and the company's reputation could collapse. Yet if Alex refused, the mechanics might turn to competitors willing to exploit the opportunity, leaving his company at a strategic disadvantage.

From a technological standpoint, the scenario described is not purely science fiction. Modern autonomous vehicles already integrate high-resolution cameras, lidar, radar, and GPS to detect, localize, and track pedestrians, vehicles, and obstacles in real time. Facial recognition systems, powered by deep neural networks such as convolutional neural networks (CNNs) or vision transformers, can identify individuals with high accuracy in controlled conditions (e.g., good lighting, a frontal view, and minimal occlusion). When combined with real-time databases (e.g., a premium client registry synced via secure cloud APIs), a vehicle could theoretically cross-reference detected faces against a list of "high-priority" individuals. Machine learning models can also be trained to estimate demographic or social attributes (e.g., age, attire, gait) as proxies for status (though this introduces significant bias and error). Decision-making algorithms, such as reinforcement learning agents or rule-based ethical governors, could then adjust trajectory planning to minimize harm to identified high-value targets, for example by steering away from them even if it increases risk to others. While current safety-critical automotive systems (e.g., ISO 26262-compliant) prioritize *minimizing total harm* and avoid identity-based logic due to legal and ethical concerns, the underlying sensing, identification, and path-planning capabilities *do exist* and what is considered harm is not universal.

## CLOSING

As Alex left the meeting, the dilemma kept his mind quite busy. Should he authorize the potential changes in the code that privilege influential clients, so that competitive advantage can be secured at the expense of fairness? Or should he refuse, and preserve ethical integrity at the risk of losing a lucrative business opportunity? On the one hand, there is a potential undiscovered market for safety packages for drivers, passengers, and even pedestrians; on the other hand, this could go wrong in many different ways… The choice he made would shape not only his career but also the moral trajectory of

autonomous systems. In an age where algorithms decide who lives and who is endangered, the trolley problem is no longer a theoretical thought experiment. It is a design decision.