Week 7: Ethics and AI

Nate Bachmeier

TIM-8150: Artificial Intelligence

November 15, 2020

Northcentral University

# Ethics and AI

Artificial intelligence is a scary black box that spreads malicious propaganda, destroys jobs, and seeks to undermind honest citizens’ values. This statement is intentionally farcical, yet it also touches on real concerns of ethical AI designs. People fear what they do not understand and use science-fiction to fill these gaps. Within those futuristic worlds, machines become the dominant species that controls every decision of an enslaved human population. However, several challenges prevent this transition of power from becoming a reality, such as intelligent systems lacks actual *intelligence* (Wildberger, 1996; Hole & Ahmad, 2019; Upchurch, 2018). Instead, organizations need to assess these tools rationally, explore applications that enhance human capabilities, and remove nondifferentiating overhead.

# Roles of Artificial Intelligence

## Role in Employment

Before 1949, digging a ditch would take hours or even days with a crew of manual workers. After the invention of the backhoe, these jobs required less time with fewer employees. From the organization’s perspective, these efficiencies translate into faster time to market at lower costs. Meanwhile, the former diggers became displaced into new roles that repairs, operate, and supervises the equipment. Each of these positions comes with supply chains of supporting requirements. For instance, it takes factories to produce the backhoe parts, each staffed with hundreds of blue-collar jobs. Cities must also build universities and technical schools to train team members that will fill these roles, further expanding the job market.

Similarly, modern businesses are actively seeking methods that reduce costs and improve efficiencies through automation. The most powerful artificial intelligence applications use machines to enhance human capabilities rather than replace them (Heer, 2019; Boire, 2017). For instance, a person can write a more profound business case than a machine; however, the same machine will have fewer grammatical errors. This dichotomy exists because humans specialize in contextualizing thought versus automation uses patterns to make predictions (Schleer et al. 2019). Many professions exist as a combination of decision-making, pattern recognition, and mechanical tasks. Expert systems address specific aspects of the job requirements; however, superseding the soft-skills that unify these role components is challenging (Huang et al., 2019).

Specific low-skilled jobs, such as bank tellers and office clerical staff, are at risk of being replaced (Hamid et al., 2017). Similarly, expert pattern matching tasks like identifying tumors in MRI (Magnetic Resonance Imaging) becomes commoditized. Given the lower entry barrier, some low-skilled workers will transition to better-paying jobs that operate those sophisticated and commoditized systems. For instance, many workers cannot access foreign markets due to language and communication limitations. Artificial intelligence can aid these in these translation scenarios while leaving control with humans.

## Role in Decision Making

Many decision-making processes can benefit from machines providing recommendations and validations. For instance, a court judge could use an intelligent system to assess how their sentencing aligns with existing norms. Perhaps the machine predicts the defendant should receive five years of probation, while a judge considers fifteen years in prison. When the validation check expresses such a difference in opinions, it could suggest that unconscious bias is taking place and warrants additional considerations. That bias either provides ammunition for appeals processes or incarcerates people unjustifiably long.

While this approach has much potential, there are concerns that professionals arbitrarily accept recommendations. However, these challenges occur everywhere that automation controls the ‘last mile’ of decision making. If the suggestion comes from a machine or peer, the person in charge of the process must be accountable for the final call. Blindly delegating control to machines is dangerous, precisely because learning algorithms being greedy, brittle, rigid, and opaque (Hole & Ahmad, 2019). Until artificial brains can rationalize abstract thought, humans must perform this task.

## Role in Manipulation

Modern censorship does not restrict free speech; instead, it increases the noise and drowns the signal (Thomas, 2019). Fundamentally, marketing campaigns and propaganda machines follow the same process of Segmentation Targeting and Positioning (STP) (Kane, 2019). Delivering on this objective requires pattern matching, content delivery, and human intuition. Automation is well-suited for these tasks and can use social media channels, like Facebook and Twitter, to connect with billions of people and manage significant portions of those interactions.

Congressional and media sources raise ethical questions around the ease of access to these capabilities for political manipulation. Unfortunately, these questions are mostly talking points rather than a call for action. Artificial intelligence comes with many abstract concepts that do not fit within the complex and opaque legal language (Guiffrida et al., 2018). For instance, machines cannot reason about their instructions, so can the courts hold *it* accountable? Perhaps the system designers should be responsible for their creations. However, the algorithms are primarily algebraic formulas controlled by end-users. Without a mechanism to define and enforce a standard operating behavior, it is impossible to expect a different outcome.

# Design Considerations

Two attempts to define this mechanism are the European Union’s Ethics Guidelines for Trustworthy AI and the OECD’s Principals of Ethical AI (EU, 2019; OECD, 2019). Both documents describe the need for artificially intelligent systems to be human-centric, transparent and explainable, robust and secure, safe, and accountable.

## Human-Centric

Robotics’s Three Law states that automation should not injure humans, ignore people’s commands, and protect their existence (Asimov, 1942). These rules lay a foundation of ideas that devices exist to cooperate and enhance humanity. Unfortunately, the machines cannot reason and therefore are slaves to their program designs. Since machines cannot devise these criteria independently, it becomes the system engineers’ responsibility to enforce these requirements. However, ensuring those behaviors exist is a matter of business priorities.

## Transparent and Explainable

## Robust and Secure

## Accountable