Engaging Science, Technology, and Society

Volume 7. Issue 1. (2020)

Inaugural Editorial

Original Research Articles

Bernal Prize Forum (Engagements)

Responses:

Responses:

Perspectives

Enchanted Determinism: Power without Responsibility in Artificial Intelligence

Alexander Campolo, University Of Chicago
Kate Crawford, New York University, Microsoft Research

Abstract

Deep learning techniques are growing in popularity within the field of artificial intelligence (AI). These approaches identify patterns in large scale datasets, and make classifications and predictions, which have been celebrated as more accurate than those of humans. But for a number of reasons, including nonlinear path from inputs to outputs, there is a dearth of theory that can explain why deep learning techniques work so well at pattern detection and prediction. Claims about "superhuman" accuracy and insight, paired with the inability to fully explain how these results are produced, form a discourse about AI that we call enchanted determinism.

To analyze enchanted determinism, we situate it within a broader epistemological diagnosis of modernity: Max Weber's theory of disenchantment. Deep learning occupies an ambiguous position in this framework. On one hand, it represents a complex form of technological calculation and prediction, phenomena Weber associated with disenchantment. On the other hand, both deep learning experts and observers deploy enchanted, magical discourses to describe these systems' uninterpretable mechanisms and counter-intuitive behavior. The combination of predictive accuracy and mysterious or unexplainable properties results in myth-making about deep learning's transcendent, superhuman capacities, especially when it is applied in social settings. We analyze how discourses of magical deep learning produce techno-optimism, drawing on case studies from game-playing, adversarial examples, and attempts to infer sexual orientation from facial images. Enchantment shields the creators of these systems from accountability while its deterministic, calculative power intensifies social processes of classification and control.

Keywords

artificial intelligence; deep learning; enchantment; Max Weber; magic; classification

1. The Return of Alchemy in Artificial Intelligence

In 1961, the School of Industrial Management at M.I.T. celebrated its centennial with a lecture series titled "Management and the Computer of the Future." At its conclusion, John McCarthy, organizer of the 1956 Dartmouth Conference that launched the field of artificial intelligence, made a memorable declaration. Against a consensus that sought to modestly define the different types of tasks that humans and computers were best suited to, McCarthy boldly argued that the differences between human and machine tasks were illusory. There were simply some complicated tasks that would take more time to be formalized and solved by machines (Greenberger 1962, 315). The brothers Hubert and Stuart Dreyfus were so struck by this assertion that they submitted additional remarks to the organizers, where they criticized McCarthy's equivalence between mind and machine. Instead, they used a critical metaphor of magic as a call for humility in AI research. Researchers like McCarthy "should run the same risks as the alchemist trying to synthesize gold from base materials: obscurity until success" (Greenberger 1962, 322).

Hubert Dreyfus later expanded these remarks into a report titled "Alchemy and Artificial Intelligence," in which he argued that the excessive techno-optimism in the early years of AI were driven by simplistic and problematic metaphors about intelligence, where the human brain was understood as analogous to a computer. These metaphors were misleading the field and being used to obscure the conceptual limitations and technical pitfalls they were encountering. Dreyfuss expanded his critical comparison between AI and magic, writing "the long range of alchemy has shown that any research which has had an early success can always be justified and continued by those who prefer adventure to patience" (Dreyfus 1965, 85). In other words, the surprising early efficacy of both alchemy and AI research served to cover over larger conceptual problems that prevented them from reaching a more respectable scientific status. The polemical comparison of AI with a premodern, protoscientific, magical practice[3] was designed to attack its scientific legitimacy and puncture the overconfidence of its proponents.

Much in AI has changed since the 1960s, including a shift from symbolic systems to the more recent focus on machine learning techniques. Over the last decade, AI has expanded as a field in academia and industry; now a small number of powerful technology corporations deploy AI systems at an international scale. In spite of these changes, contemporary researchers, including leaders in the field, have once again begun to describe the latest deep learning techniques as magical. In a recent interview, the computer scientist Stuart J. Russell reprises this theme:

> We are just beginning now to get some theoretical understanding of when and why the deep learning hypothesis is correct, but to a large extent, it's still a kind of magic, because

Since Dreyfus's early polemical use of alchemy, research in the history of early modern science has problematized overly sharp distinctions between conceptions of alchemy and the modern sciences. This literature instead emphasizes continuities between alchemy, magic, and chemistry as well as the great interest of historical figures like Isaac Newton and Robert Boyle in alchemical practices (Principe and Newman 2001).

Invoking magic, Russell suggests that deep neural networks can interpret real world phenomena like images, producing effective predictions without theoretical understanding of why this is so. Russell is not the only figure in deep learning to come to this conclusion. Other experts are more critical, adopting Dreyfus's polemical tone. François Chollet of Google recently characterized ad-hoc approaches to modifying learning algorithms as "folklore and magic spells" (Edwards and Edwards 2018). Ali Rahimi, also of Google, invoked "alchemy" directly to describe the lack of understanding of why certain models work (Hutson 2018). A recent event at Princeton University's Institute for Advanced Study convened some of the field's top researchers, broaching the question directly in its title: "Deep Learning: Alchemy or Science?" (Arora 2019).

This article analyzes the significance of this pattern, a discursive event that connects deep learning and magic in historically specific ways (Foucault 1972, 27). What are the features of contemporary deep learning systems and their social applications that have led them to be characterized this way, and what effects do such statements produce?

Many experts, such as Russell, imply that it is only a matter of time until we gain theoretical understanding of deep learning's predictive efficacy. Perhaps. But the discourse of enchantment operates in the present, shaping both social perceptions of these systems and the practices of their designers. It is not reducible to marketing hype or journalistic license, although both of these may reinforce popular perceptions about magic. In fact, this discourse is significant precisely because discussion of magic moves across a wide range of social positions, from experts in the field, to its critics, and to a wider public that is beginning to be exposed to deep learning's social applications.

We term this ensemble enchanted determinism: a discourse that presents deep learning techniques as magical, outside the scope of present scientific knowledge, yet also deterministic, in that deep learning systems can nonetheless detect patterns that give unprecedented access to people's identities, emotions and social character. These systems become deterministic when they are deployed unilaterally in critical social areas, from healthcare to the criminal justice system, creating ever more granular distinctions, relations, and hierarchies that are outside of political or civic processes, with consequences that even their designers may not fully understand or control. New problems arise when the lived effects of social prediction and categorization are unknown to their makers and unaccountable to those who are disadvantaged by them when applied in the world. The application of these systems threatens not only legal due process (Citron and Pasquale 2014) but also more expansive forms of political contestation, and social agency, while simultaneously distancing AI designers and the corporations that employ them from ethical responsibility and legal liability.

2. Enchantment and Disenchantment in Deep Learning

It is often the case that new technologies are presented as magical, and contemporary forms of deep learning are no exception. A number of scholars have shown how those with an interest in marketing and profiting from AI benefit from this association. M.C. Elish and danah boyd use the idea of magic to analyze "the manufacturing of hype and promise," which allows businesses to "produce a rhetoric around these technologies that extends far past the current methodological capabilities" (2018, 58). Similarly, Emmanuel Moss and Friederike Schüür show how mythic metaphors build an understanding of machine learning systems as "superhuman" in ways that implicitly separate them from the human capabilities and practices needed for their implementation (2018, 278). There is no doubt that discourses of magic contribute to the intense contemporary hype around AI in this wider sense.

The discourse of enchanted determinism goes beyond marketing or press hype that covers over technological shortcomings of deep learning and its social applications. Instead it operates when these systems succeed, at least according to the narrow engineering criteria selected by their creators, when magical mystery and technical mastery curiously work together.

Max Weber's theory of disenchantment allows us to draw out epistemological and political issues at play in the social application of deep learning systems.[4] Disenchantment—a more literal translation of his German phrase "Entzauberung" would be "de-magification"—is an epochal diagnosis of Western modernity, encompassing a widespread decline in mystical or religious forces[5] and their replacement by processes of "rationalization and intellectualization" (Weber 1946, 139). This social process encompasses the rise of modern science, whose concepts and experiments contrast with magical ways of understanding the world.[6] Disenchantment

4 We are not the only scholars to have recently returned to classical Weberian concepts to analyze contemporary technological developments. Morgan Ames (2014, 2015) has used the Weberian notion of charisma—often associated with magic in his sociology of religion (Riesebrodt 1999)—to analyze the ways that technology operates ideologically to both promise solutions to social problems while simultaneously working to conserve an existing social order.

5 While there is an extensive Weberian literature dedicated to the broader relationship between science and religion in modernity (Asad 2003, Taylor 2007, Scott 2017), our purpose is different. We are interested in when and why themes of enchantment and disenchantment recur in specific historical situations. In other words, our argument is not that enchantment is a useful analytic because "we have never been modern" (Latour 1993) or even that have never been disenchanted, as some of Weber's critics have recently suggested (Bennett 2001, Josephson-Storm 2017). Instead the concept of enchantment gives allows us to grasp how new technologies challenge models of causality, mastery, and the social itself in historically and culturally specific ways.

6 Magic, of course, is a topic with a rich history in the social sciences, whose study predates Weber. Anthropologists have been particularly attentive to the use of magic in human societies. Many of the discipline's most influential early practitioners, from E.B. Tylor to Sir James Frazer, conceived of magic in terms similar to those with which Dreyfus characterized alchemy—as essentially mistaken premodern practices that science would replace on a historical path of modernization. Subsequent work in anthropology has taken ethnographic data on magic more seriously in order to understand models of causality (Winkelman 1982).

Engaging Science, Technology, and Society

Volume 7. Issue 1. (2020)

Inaugural Editorial

Original Research Articles

Bernal Prize Forum (Engagements)

Perspectives

Enchanted Determinism: Power without Responsibility in Artificial Intelligence

Alexander Campolo, University Of Chicago
Kate Crawford, New York University, Microsoft Research

Abstract

Deep learning techniques are growing in popularity within the field of artificial intelligence (AI). These approaches identify patterns in large scale datasets, and make classifications and predictions, which have been celebrated as more accurate than those of humans. But for a number of reasons, including nonlinear path from inputs to outputs, there is a dearth of theory that can explain why deep learning techniques work so well at pattern detection and prediction. Claims about "superhuman" accuracy and insight, paired with the inability to fully explain how these results are produced, form a discourse about AI that we call enchanted determinism.

To analyze enchanted determinism, we situate it within a broader epistemological diagnosis of modernity: Max Weber's theory of disenchantment. Deep learning occupies an ambiguous position in this framework. On one hand, it represents a complex form of technological calculation and prediction, phenomena Weber associated with disenchantment. On the other hand, both deep learning experts and observers deploy enchanted, magical discourses to describe these systems' uninterpretable mechanisms and counter-intuitive behavior. The combination of predictive accuracy and mysterious or unexplainable properties results in myth- making about deep learning's transcendent, superhuman capacities, especially when it is applied in social settings. We analyze how discourses of magical deep learning produce techno-optimism, drawing on case studies from game-playing, adversarial examples, and attempts to infer sexual orientation from facial images. Enchantment shields the creators of these systems from accountability while its deterministic, calculative power intensifies social processes of classification and control.

Keywords

artificial intelligence; deep learning; enchantment; Max Weber; magic; classification

1. The Return of Alchemy in Artificial Intelligence

In 1961, the School of Industrial Management at M.I.T. celebrated its centennial with a lecture series titled "Management and the Computer of the Future." At its conclusion, John McCarthy, organizer of the 1956 Dartmouth Conference that launched the field of artificial intelligence, made a memorable declaration. Against a consensus that sought to modestly define the different types of tasks that humans and computers were best suited to, McCarthy boldly argued that the differences between human and machine tasks were illusory. There were simply some complicated tasks that would take more time to be formalized and solved by machines (Greenberger 1962, 315). The brothers Hubert and Stuart Dreyfus were so struck by this assertion that they submitted additional remarks to the organizers, where they criticized McCarthy's equivalence between mind and machine. Instead, they used a critical metaphor of magic as a call for humility in AI research. Researchers like McCarthy "should run the same risks as the alchemist trying to synthesize gold from base materials: obscurity until success" (Greenberger 1962, 322).

Hubert Dreyfus later expanded these remarks into a report titled "Alchemy and Artificial Intelligence," in which he argued that the excessive techno-optimism in the early years of AI were driven by simplistic and problematic metaphors about intelligence, where the human brain was understood as analogous to a computer. These metaphors were misleading the field and being used to obscure the conceptual limitations and technical pitfalls they were encountering. Dreyfuss expanded his critical comparison between AI and magic, writing "the long range of alchemy has shown that any research which has had an early success can always be justified and continued by those who prefer adventure to patience" (Dreyfus 1965, 85). In other words, the surprising early efficacy of both alchemy and AI research served to cover over larger conceptual problems that prevented them from reaching a more respectable scientific status. The polemical comparison of AI with a premodern, protoscientific, magical practice3 was designed to attack its scientific legitimacy and puncture the overconfidence of its proponents.

Much in AI has changed since the 1960s, including a shift from symbolic systems to the more recent focus on machine learning techniques. Over the last decade, AI has expanded as a field in academia and industry; now a small number of powerful technology corporations deploy AI systems at an international scale. In spite of these changes, contemporary researchers, including leaders in the field, have once again begun to describe the latest deep learning techniques as magical. In a recent interview, the computer scientist Stuart J. Russell reprises this theme:

> We are just beginning now to get some theoretical understanding of when and why the deep learning hypothesis is correct, but to a large extent, it's still a kind of magic, because

Since Dreyfus's early polemical use of alchemy, research in the history of early modern science has problematized overly sharp distinctions between conceptions of alchemy and the modern sciences. This literature instead emphasizes continuities between alchemy, magic, and chemistry as well as the great interest of historical figures like Isaac Newton and Robert Boyle in alchemical practices (Principe and Newman 2001).

Invoking magic, Russell suggests that deep neural networks can interpret real world phenomena like images, producing effective predictions without theoretical understanding of why this is so. Russell is not the only figure in deep learning to come to this conclusion. Other experts are more critical, adopting Dreyfus's polemical tone. François Chollet of Google recently characterized ad-hoc approaches to modifying learning algorithms as "folklore and magic spells" (Edwards and Edwards 2018). Ali Rahimi, also of Google, invoked "alchemy" directly to describe the lack of understanding of why certain models work (Hutson 2018). A recent event at Princeton University's Institute for Advanced Study convened some of the field's top researchers, broaching the question directly in its title: "Deep Learning: Alchemy or Science?" (Arora 2019).

This article analyzes the significance of this pattern, a discursive event that connects deep learning and magic in historically specific ways (Foucault 1972, 27). What are the features of contemporary deep learning systems and their social applications that have led them to be characterized this way, and what effects do such statements produce?

Many experts, such as Russell, imply that it is only a matter of time until we gain theoretical understanding of deep learning's predictive efficacy. Perhaps. But the discourse of enchantment operates in the present, shaping both social perceptions of these systems and the practices of their designers. It is not reducible to marketing hype or journalistic license, although both of these may reinforce popular perceptions about magic. In fact, this discourse is significant precisely because discussion of magic moves across a wide range of social positions, from experts in the field, to its critics, and to a wider public that is beginning to be exposed to deep learning's social applications.

We term this ensemble enchanted determinism: a discourse that presents deep learning techniques as magical, outside the scope of present scientific knowledge, yet also deterministic, in that deep learning systems can nonetheless detect patterns that give unprecedented access to people's identities, emotions and social character. These systems become deterministic when they are deployed unilaterally in critical social areas, from healthcare to the criminal justice system, creating ever more granular distinctions, relations, and hierarchies that are outside of political or civic processes, with consequences that even their designers may not fully understand or control. New problems arise when the lived effects of social prediction and categorization are unknown to their makers and unaccountable to those who are disadvantaged by them when applied in the world. The application of these systems threatens not only legal due process (Citron and Pasquale 2014) but also more expansive forms of political contestation, and social agency, while simultaneously distancing AI designers and the corporations that employ them from ethical responsibility and legal liability.

2. Enchantment and Disenchantment in Deep Learning

It is often the case that new technologies are presented as magical, and contemporary forms of deep learning are no exception. A number of scholars have shown how those with an interest in marketing and profiting from AI benefit from this association. M.C. Elish and danah boyd use the idea of magic to analyze "the manufacturing of hype and promise," which allows businesses to "produce a rhetoric around these technologies that extends far past the current methodological capabilities" (2018, 58). Similarly, Emmanuel Moss and Friederike Schüür show how mythic metaphors build an understanding of machine learning systems as "superhuman" in ways that implicitly separate them from the human capabilities and practices needed for their implementation (2018, 278). There is no doubt that discourses of magic contribute to the intense contemporary hype around AI in this wider sense.

The discourse of enchanted determinism goes beyond marketing or press hype that covers over technological shortcomings of deep learning and its social applications. Instead it operates when these systems succeed, at least according to the narrow engineering criteria selected by their creators, when magical mystery and technical mastery curiously work together.

Max Weber's theory of disenchantment allows us to draw out epistemological and political issues at play in the social application of deep learning systems.4 Disenchantment—a more literal translation of his German phrase "Entzauberung" would be "de-magification"—is an epochal diagnosis of Western modernity, encompassing a widespread decline in mystical or religious forces5 and their replacement by processes of "rationalization and intellectualization" (Weber 1946, 139). This social process encompasses the rise of modern science, whose concepts and experiments contrast with magical ways of understanding the world.6 Disenchantment

4 We are not the only scholars to have recently returned to classical Weberian concepts to analyze contemporary technological developments. Morgan Ames (2014, 2015) has used the Weberian notion of charisma—often associated with magic in his sociology of religion (Riesebrodt 1999)—to analyze the ways that technology operates ideologically to both promise solutions to social problems while simultaneously working to conserve an existing social order.

5 While there is an extensive Weberian literature dedicated to the broader relationship between science and religion in modernity (Asad 2003, Taylor 2007, Scott 2017), our purpose is different. We are interested in when and why themes of enchantment and disenchantment recur in specific historical situations. In other words, our argument is not that enchantment is a useful analytic because "we have never been modern" (Latour 1993) or even that have never been disenchanted, as some of Weber's critics have recently suggested (Bennett 2001, Josephson-Storm 2017). Instead the concept of enchantment gives allows us to grasp how new technologies challenge models of causality, mastery, and the social itself in historically and culturally specific ways.

6 Magic, of course, is a topic with a rich history in the social sciences, whose study predates Weber. Anthropologists have been particularly attentive to the use of magic in human societies. Many of the discipline's most influential early practitioners, from E.B. Tylor to Sir James Frazer, conceived of magic in terms similar to those with which Dreyfus characterized alchemy—as essentially mistaken premodern practices that science would replace on a historical path of modernization. Subsequent work in anthropology has taken ethnographic data on magic more seriously in order to understand models of causality (Winkelman 1982).