Skin in the Game: Modulate AI and Addressing the Legal and Ethical Challenges of Voice Skin Technology

Citation

Gordon, Rachel, and Ryan Budish. 2020. "Skin in the Game: Modulate AI and Addressing the Legal and Ethical Challenges of Voice Skin Technology." BKC Policy Practice on AI, Berkman Klein Center for Internet & Society.

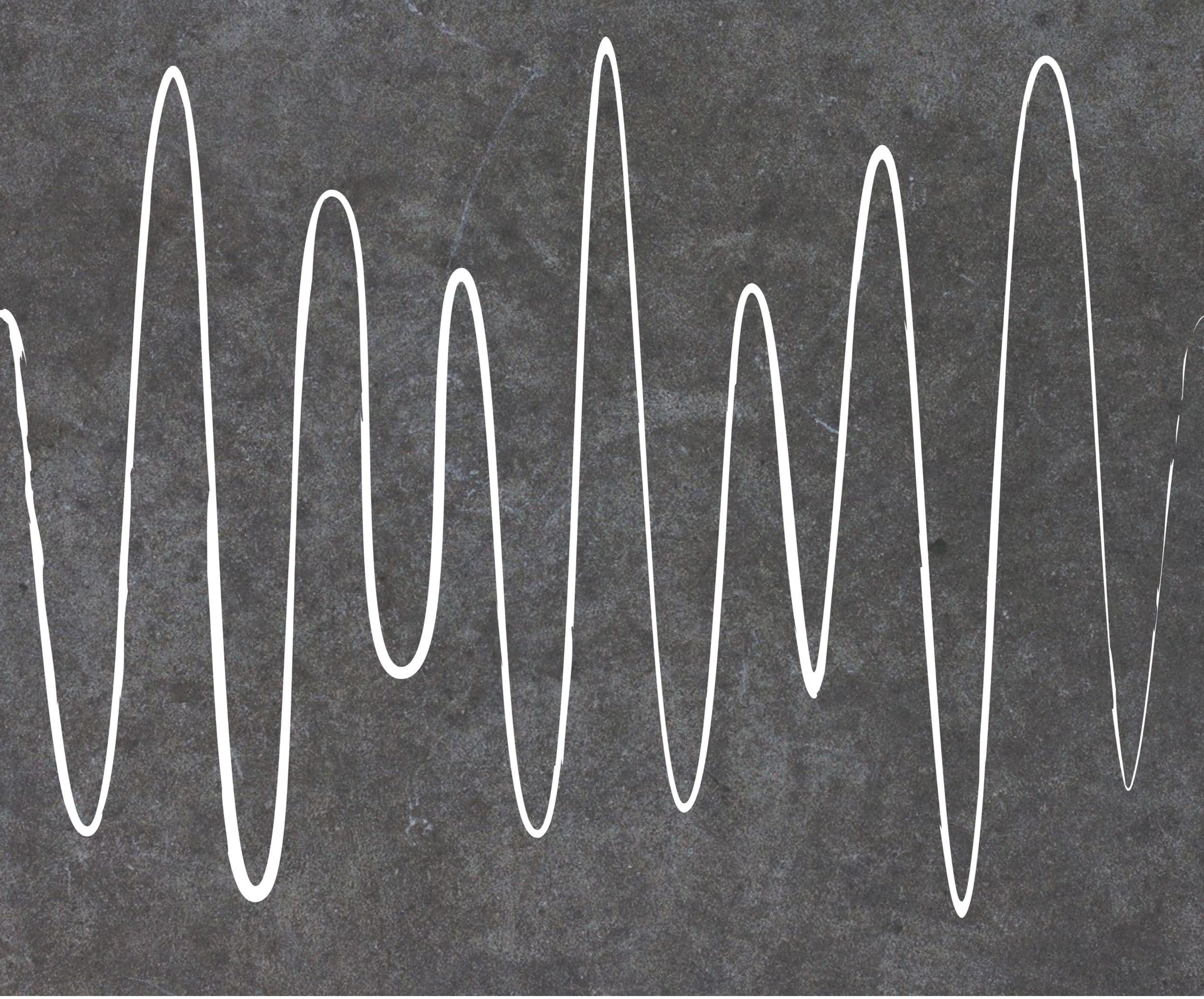
SKIN IN THE GAME: A TOOLKIT EXPLORING THE LEGAL AND ETHICAL CHALLENGES OF VOICE SKIN TECHNOLOGY

The following is an educational toolkit from the BKC Policy Practice on AI that includes a case study, a teaching note, and a background primer. Collectively, they comprise a toolkit that can illuminate some of the challenges in moving from AI principles to practice.

Toolkit Components

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY
AT HARVARD UNIVERSITY

HARVARD LAW SCHOOL | The Case Studies



SKIN IN THE GAME: MODULATE AI AND ADDRESSING THE LEGAL AND ETHICAL CHALLENGES OF VOICE SKIN TECHNOLOGY

RACHEL GORDON • RYAN BUDISH

CYBER.HARVARD.EDU

HARVARD LAW SCHOOL | The Case Studies

This case study is produced by the BKC Policy Practice: Artificial Intelligence (AI) at the Berkman Klein Center for Internet & Society at Harvard University. BKC Policy Practice: AI is a public interest-oriented program that helps governmental, nonprofit, and private sector organizations implement AI best practices and turn AI principles into operational realities. The AI policy practice connects decision makers with faculty, students, and experts from across the BKC community.

Since 2018, the Berkman Klein Center has hosted workshops with a variety of public and private sector organizations, creating a space for learning, knowledge-sharing, and capacity-building in which participating organizations work with small, agile teams composed of faculty, staff, students, and outside expert collaborators to identify key problems and create actionable outputs. Outputs range from policy frameworks to open educational resources and are developed with an eye towards inclusion, replicability, and broad usefulness. This case emerged from these workshops.

Written by Rachel Gordon, Research Associate, Teaching Learning and Curriculum Solutions, Harvard Law School Library, and Ryan Budish, Assistant Research Director, Berkman Klein Center for Internet & Society at Harvard University, this case was developed as a basis for discussion in educational and training environments. It is not intended to serve as an endorsement, source of primary data, legal advice, or illustrative example of effective or ineffective practices.

The Case Studies Program supports a wide range of case development projects throughout Harvard Law School. Faculty at HLS are in the process of creating innovative, experiential materials for the legal classroom, and the Case Studies Program works with them to conceive, develop, edit, and publish their classroom materials. To find more HLS Case Studies visit: https://casestudies.law.harvard.edu.

This case is open source and available on the Berkman Klein Center website at http://cyber.harvard.edu/publication/2020/modulate-case-study. For more information about the Berkman Klein Center's Open-Access Policy, please visit https://cyber.harvard.edu/node/open-access-policy.

BKC Policy Practice: AI is supported by funding from the Ethics and Governance of AI Fund.

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY
AT HARVARD UNIVERSITY

HARVARD
LAW SCHOOL

The Case Studies

CONTENTS

Introduction

In 2017, Carter Huffman and Mike Pappas co-founded Modulate. The fellow MIT alums and friends created Modulate to commercialize the unique voice-generation technology that Huffman had invented after working at NASA's Jet Propulsion Laboratory. By innovatively applying concepts from artificial intelligence systems called Generative Adversarial Networks (GANs),[1] the two men had created a novel—even transformative—approach to make one voice sound like another in real time. Now, in February 2019, they were excited—they had just secured US$2 million in venture capital funding enabling them to launch their technology commercially. Despite their passion and these venture funds, the reality remained that they were still a small start-up with a total of three employees working out of a shared incubator space located in Cambridge, Massachusetts. Although Modulate had extremely limited human and financial resources, Huffman and Pappas wanted to ensure that this technology, with its ability to match the timbre[2] of almost any individual on the planet, would not be misused. How could they simultaneously push Modulate forward, maintain both its technological and competitive edges and make their investors happy while also upholding a code of ethics? For a tiny company like Modulate, what did this look like?

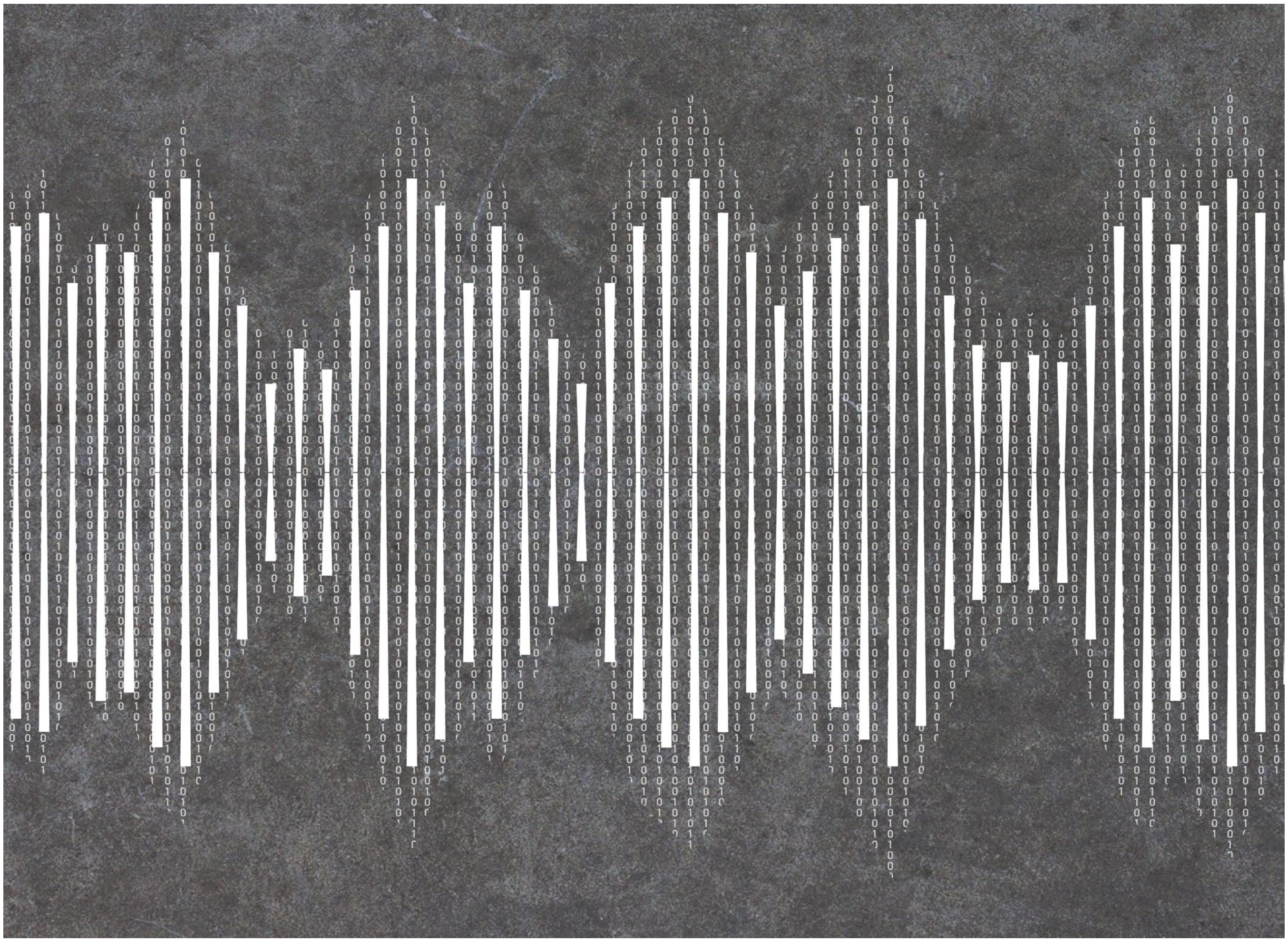
A Technology Looking for a Commercial Application

Even before Modulate secured its venture capital funding, Huffman and Pappas had been so enthusiastic about the possibilities of their new technology that they quit their jobs in May 2018 to work on Modulate full time. They had already incorporated in August 2017, and by September 2018, felt that they had enough direction to start shopping the Modulate technology to possible funders. Huffman noted, "This started as a technical challenge. We thought that we ought to be able to do this. After about [a] year and half, the technology was good enough. It sounded like a human voice. Then we began to wonder, 'What new applications can we enable with these new kinds of tools? What kind of business could we start?'"

Throughout this period, Huffman and Pappas considered both the commercial applications and the ethical implications for their technology. Creating a direct-to-consumer downloadable app that allowed people to change their voices seemed like an obvious choice, but the men thought it would be too susceptible to abuse. Huffman remarked, "People encouraged us to release a phone app that could make prank calls—similar to those which existed using old school voice changers, which while not convincing, can still hide your voice. We were against doing that." Pappas emphasized, "We wanted to limit the use of our technology so that it would not be used for harm." Some malicious uses they feared included the misuse of their voice skins to imitate a politician, to create misleading news stories, or to imitate a friend or family member in order to commit fraud. However, to demonstrate

[1] For more background about GANs especially within a legal context see the accompanying primer, *Background Primer: "Skin in the Game: Modulate AI and Addressing the Legal and Ethical Challenges of Voice Skin Technology"*, available at http://cyber.harvard.edu/publication/2020/modulate-case-study

[2] Modulate's technology did not directly match all components of one's speech such as cadence, speech style or emotiveness but it did imitate timbre, the quality and tone that makes a voice sound unique.



they could replicate the voice of a celebrity,the website had a clip of President Barack Obama "speaking" that was almost indistinguishable from an actual recording of President Obama speaking.[3] Neither the Obama voice clip, nor any others like it were available to use as voice skins.

Huffman and Pappas met regularly with MIT Venture Mentoring Service (VMS) volunteers who, according to Pappas, "gave us guidance on conducting solid market research and thinking about what actually counts as evidence for a market's existence." The men felt like MIT VMS had offered critical feedback which reinforced to the two men that they needed a specific commercial approach to leverage their powerful technology. Pappas remembered, "We had spent months testing different commercial approaches, and a pitch to VMS was the inflection point where we realized we needed to make a decision." Concerned about the potential misuse of their technology, Huffman and Pappas decided not to make their technology available to consumers, and instead sell only to other businesses. Eventually, based on the market research and a joint love of gaming experiences and video game connections, Huffman and Pappas decided to focus on the video game industry.

The possibilities within the video game industry excited the pair. The industry was typically receptive to new technology and people within the industry had even sought Huffman and Pappas out to learn about voice skins. Huffman explained, "In a game people already inhabit a virtual avatar. Now they can have a voice chosen

[3] To listen to the demo go to https://modulate.ai/. Accessed November 2019.

by them to match." Huffman and Pappas felt that the end users—game players—would be intrigued enough to purchase voice skins, but most importantly, that voice skins would amplify the experience just as the best existing features for customizing an avatar did."[4] They believed players would see voice skins as an exciting and novel way to enhance their online personas. Pappas explained,

> Our voice skins offer a method of expression for players. Imagine if someone was a big bulky avatar. You might choose a deep commanding voice for your avatar. Or you might play with expectations and instead of picking a deep voice you might pick a high pitched squeaky voice. People already mix and match all the time in the game space. That is the freedom that we want to give players.

Pappas and Huffman were excited about the possibilities provided by the video game industry, but wondered how to best tap into this market. One option was to sell voice skins directly to players. This approach would allow Modulate to create new voices, add new features, and improve the technology overall. However, they both felt strongly that this route would make Modulate's technology ***too accessible*** with a high potential for misuse. Instead, Huffman and Pappas decided that Modulate would work directly with video game companies. In this model, Modulate would work directly with video game designers and publishers to create a limited number of specific voices. These voice skins would be available to players as a menu of in-game options that players could select or buy for their in-game avatars. Players would not be able to create any voice they wanted, but would be limited to the voices created by Modulate that the game designers had selected for the game. Huffman commented, "We felt like if a business was our customer, we would have more control over how our technology got used."

Huffman and Pappas continued to improve Modulate's technology and develop a commercial strategy while meeting with venture capitalists. During their pitches Huffman and Pappas explained that they had built a watermarking capability specifically to make it harder for users to deceive others about their identities. VC reactions to this feature varied. A few viewed it as unnecessary; others thought it was a smart proactive move to deal with voice fraud and malicious impersonation seeing it as a feature that distinguished Modulate from competitors. Ultimately, the two companies that gave Modulate its seed funding valued Modulate's proactive approach to handling hypothetical abuse issues. Huffman and Pappas felt lucky to have found investors who supported spending some of their limited time and resources to try to prevent misuse of their technology.

Creating Value while Preserving Values

Lawful Use of Technology

Huffman and Pappas wanted to ensure that their voices were unbiased and limit the technology's potential for misuse. But their most significant and pressing concern was staying within the bounds of the law. Although their technology presented several legal questions, one important need was proactively developing a strategy to ensure that a synthetic voice skin did not sound too similar to a famous personality. This was not an issue that Modulate faced yet, but they worried that it could happen as they built out their library of synthetic

[4] For example, developers offered the popular video game Fortnite for free yet still had 2018 revenues of $2.4 billion from "micro-transactions." These occurred when Fortnite players bought accessories for their game avatars within the Fortnite platform. To read more see, Ganti, A. (2019, December 4). How Does Fortnite Make Money: Monetizing Exclusivity? Retrieved December 20, 2019, from https://www.investopedia.com/tech/how-does-fortnite-make-money/.

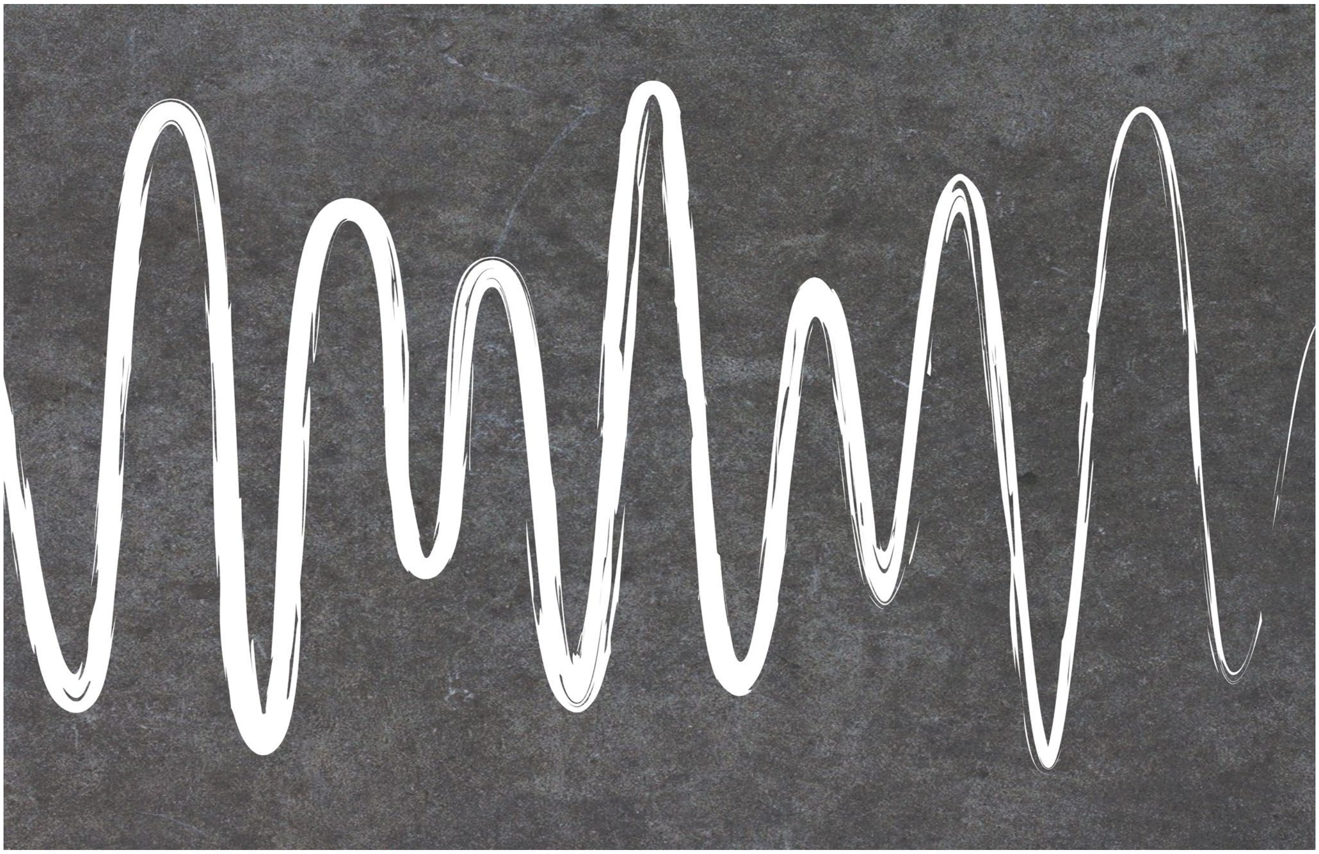
voice skins. To sidestep this issue, before creating a voice skin for a game maker, Modulate required that customers demonstrate they had the rights to the resulting voice. But as Pappas explained, “Even that is ambiguous as in what exactly are the rights to a voice in certain circumstances? What would happen if a game developer hired an actor to provide the voice for a voice skin, but that person’s voice sounded like a celebrity’s voice?” Pappas continued, “Coincidences where people’s voices sound similar to other peoples will become more frequent or at least easier to ‘get close enough to’ as we create more voices.” Hence controlling voice creation by building their own in-house library and offering customers a catalog of voices from which to choose was a strategic effort to avoid potential lawsuits for copyright infringement and privacy violations.

Accounting for and Managing Bias

With a business model in place and the most obvious legal issues seemingly avoided, Modulate needed product—voice skins—to sell. Initially, the men turned to a public dataset for Modulate’s voice samples. This dataset, however, turned out to be too limited—both in terms of diversity and emotiveness—as it contained mainly British-English speakers reading from Wikipedia. The men tried to seek out a diverse set of communities to build their dataset by hiring voice actors to record in the company’s studio. While this training data was more diverse, it drew only on the voice actors available to them. Pappas wondered, “Even if we’re actively soliciting voice actors with different backgrounds, we’re still limited to those in the Boston area. What’s more, there’s no well-defined set of all voices, so it’s impossible to know if we’ve covered all our bases without simply putting the technology out there. So we’re constantly challenging ourselves to think of new kinds of users or use cases which would require new data, and then collecting that data as proactively as we can.” (Later, they also recorded their own voices to add to the dataset.)

> **“Coincidences where people’s voices sound similar to other peoples will become more frequent or at least easier to ‘get close enough to’ as we create more voices.”**

Building their own voice library gave Modulate more control and choice, but also made Huffman and Pappas acutely aware that they needed to offer clients a wide range of voice options. At the same time, the men could not build a library of every possible voice—how could they be sure that the voices they developed were the ones customers would want? To get around this, Huffman and Pappas worked with game developers to create specific kinds of voices. The game industry, however, had a long history of bias and discrimination, particularly along gender lines. Already, a couple of potential customers had told the men that “a lot of gamers want to sound sexier.” Pappas said, “That is not necessarily inherently bad,” but he acknowledged that “there is probably a bit of social biases in terms of what qualifies as sexy.” If customers requested only certain kinds of voices—sexy, male, white, Western, etc.—the customers’ biases would then become biases in Modulate’s library. Moreover, because GANs got better over time, if Modulate made only certain kinds of voices, the system would become better at making those voices,



and relatively worse at other voices. Pappas acknowledged, “It is hard to say where to draw the line. I think that it would be okay if the voice skin matched the existing game so long as it wasn’t fundamentally broken or people were trying to prevent the game’s existence. Based on our model and technology, we’ll have the opportunity to act as advisors and give at least some input.”

There were technical challenges, too. Machine learning systems need to be trained on huge data sets; Modulate’s system was no different, requiring significant amounts of voice recordings. To ensure that their offerings provided a range of distinct and unique sounding voices, they needed training data from many different sounding individuals (and/or actors) reading scripts. Model training with voice data, however, was only part of the process. Once the voice skin was created and made available in a game, the model needed to take the gamer’s voice and apply the voice skin to create a new voice. Would the Modulate technology work equally well regardless of the speaker? Would the synthetic voice work as well on someone with a heavy accent as it would on someone whose accent more closely resembled Pappas, Huffman, and the people who helped test the technology?

Responsibility Limits—What Could Modulate Do? What Should It Do?

As Huffman and Pappas worked on developing their voice data set, they wondered about voices that might be unobjectionable in certain contexts but would be problematic in others. For example, would an Adolph Hitler voice skin be acceptable in a World War II game? Would allowing players to act as a different gender or race be a positive, eye-opening experience, or would it simply put minorities at risk of cultural appropriation? Even more problematic was the creation of voice skins that sounded like children, which invited a whole host of challenges around safety. If they made childlike voices, could adults misuse those voices to engage in inappropriate sexual dialogue with a child playing a game?

Modulate did not have a set policy to deal with these issues but instead took a case-by-case approach. Given the small size of the company, when an issue like this came up, the entire team could discuss the customer's request to determine an answer. But this approach would not scale, and Modulate had no process or structure in place to navigate these questions. As Pappas noted, "The only hard line now is a customer needs to show they have the rights for the voice." Yet the pair understood that these issues were multi-layered and complex. Pappas described Modulate's dilemma, "We've architected it so we will have the right to decide about each new voice. But the follow-on question is, 'How do we make that decision?' The first part is technical but then how do we ensure that whatever we are doing presently yields not just good outcomes now but in six months and even in 10 years as the landscape evolves?" Huffman interjected, "What happens too if we make a mistake?"

In the case of children's voices, Huffman and Pappas had not reached a general decision on whether those voices should be included in a voice skin library. They had planned to record a child's voice for the library but child labor laws actually made that too complicated, so they abandoned the effort. Nonetheless, the problem remained because they could still record an adult who sounded like a child and make that recording available in a voice skin library. For the moment, the two decided it was easier to not have children's voices available but, if asked, would carefully consider the use case before taking that step. Huffman elaborated, "For example, in some scenarios allowing a child to sound like another's child voice might be appropriate, whereas allowing an adult to sound like a child might not be."

Similarly, although Huffman and Pappas believed in the importance of transparency, they struggled with the question of whether it was their responsibility or their customers' responsibility to let players know when someone is using a voice skin. According to the current business model, it was up to the game designers to decide whether and how to inform game players that voice skins were in use. Huffman and Pappas chose to encourage certain norms, or "voice skin etiquette," rather than include announcements to alert people that the voice skins they were hearing were fakes. Modulate could require that their customers disclose the use of voice skins, but they worried that constantly identifying a voice as "made-up" would disrupt the immersive experience that games sought to create. Pappas explained, "It is the platform's choice to disclose actively if a player is using a voice skin. The risk is that if the voice skin use gets disclosed, it could ruin the immersion experience lessening the value of using the voice skin. If you are constantly reminded that an elf isn't really an elf it could negatively affect your experience."

It was still not entirely clear to Huffman and Pappas how much responsibility Modulate bore versus the games using their technology. What would Modulate do if any of their customers' customers misused the technology? Pappas, noting that a disguised voice enables one to hide one's identity, admitted, "There is a general question around how much it is our responsibility to provide only the appropriate set of voices versus the platform's responsibility if people are using voice skins to harass and mistreat others."

Creating an Ethical Culture

Moreover, the men had continued to think about making ethical practices and decisions a part of Modulate's culture and identity. For internal use, they developed a loose conceptual framework. Pappas explained, "There are two main vectors

> **"There are two main vectors of ethics for us. First, how much can we limit the use of our own technology? The second piece is, 'What is the context and expectations of the person hearing one of our voice skins?"**

of ethics for us. First, how much can we limit the use of our own technology? We want to ensure someone isn't using it to make a prank call in the same way that we won't put something out in the market that allows anyone to design an arbitrary voice. We want to know how our technology is used and hold our customers accountable." He went on, "The second piece is, 'What is the context and expectations of the person hearing one of our voice skins?' This is a really clear distinction. In gaming, people are accustomed to skins use in an artificial world." As time allowed, Pappas had begun developing a staff ethics handbook. Given Modulate's small size, Pappas commented, "Ethical decision making is part of our interview process but we don't have explicit directives for specific situations. As a team of four we have gone into deep detail during conversations. We want to ensure that we are not missing anything and have all the right structures to avoid something."

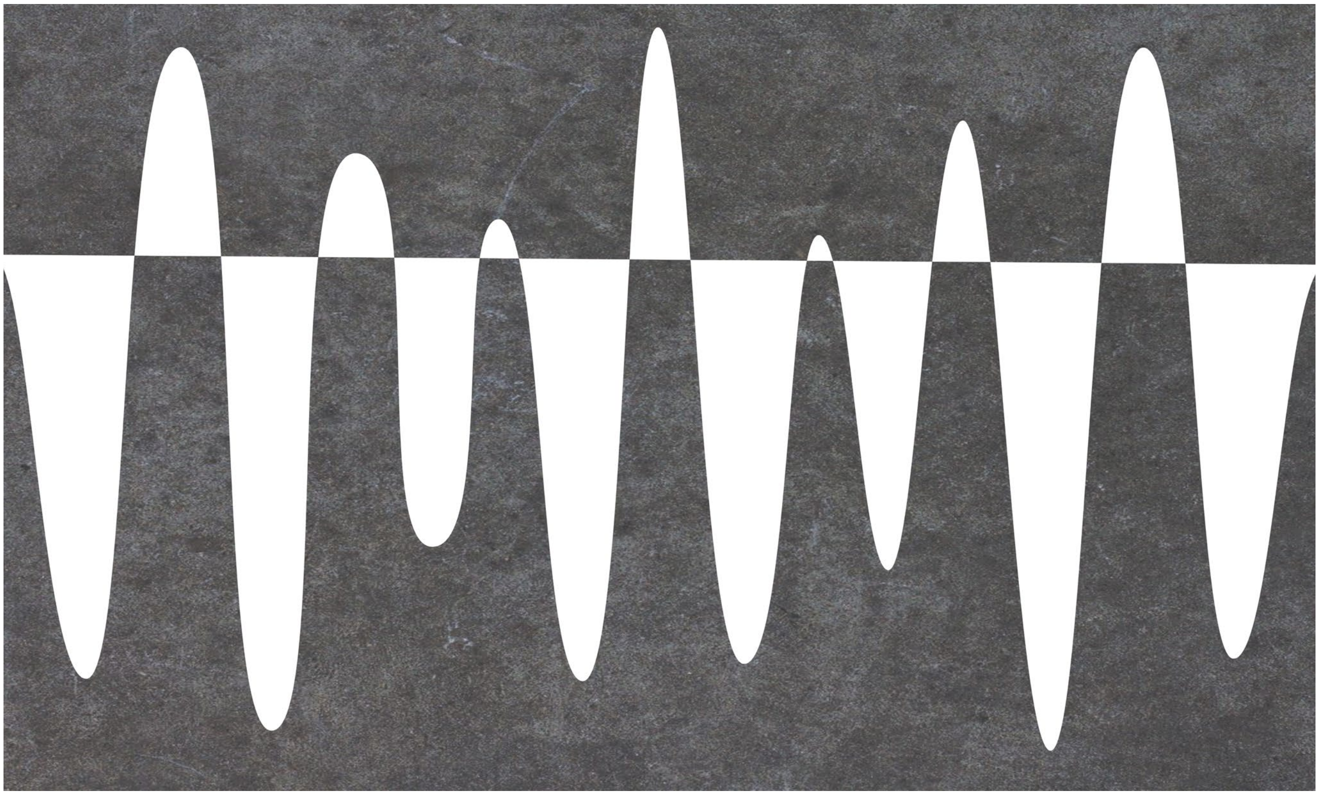
To safeguard against malicious use and to signal to the market the importance of ethical use of its technology, Modulate had developed a watermarking capability making it possible to identify a Modulate created voice. Pappas acknowledged, "There's no such thing as a perfect solution, but including watermark elements allowed us to drastically reduce the number of ways our technology could be misused, while at the same time increased the technical skill required to apply our voice skins to a malicious purpose." The company also had an Ethics FAQ section posted on its website. (See **Exhibit 1** to read the FAQ section.) An added bonus was that the Ethics FAQ section had proven to be an effective recruiting strategy. Several job seekers had reached out to Modulate after reading the ethics section.

What Lay Ahead for Modulate?

When Modulate launched, it was the only product of its kind, but Huffman and Pappas knew that they could not take their technical advantage for granted.[5] As Huffman pointed out, "We're not so naïve as to think that no one else could do this. People will try to catch up." Huffman and Pappas struggled with how to capitalize on their first-to-market position while also maintaining their position as good actors, but that came with a cost. As a small team with limited resources, every minute and every dollar spent trying to prevent misuse of their technology was time and money not spent developing new features or finding new customers.

At the end of 2019, they had a few customers, all of whom were still developing their games, meaning "voice skins" had not yet been truly let loose into the wild. Thus Huffman and Pappas had to help their existing customers but also focus on customer acquisition. They faced a quandary. Handicapping their technology too much would mean

[5] As of December 2019 Modulate did not have any competitors who used machine learning to offer voice skins in the way that Modulate did.



losing their competitive advantage making them less attractive to potential customers. Moreover, they had a fiduciary responsibility to their investors. Pappas explained, “We have an obligation to try to succeed. We have the first mover advantage but even if we do everything right, there is still a risk that someone willing to compromise their ethics could overtake us. So perhaps it makes sense to have less onerous standards and have ones that don’t feel too difficult to meet so that we have a little bit better overall ecosystem.”

Huffman and Pappas strongly believed that the positive impact of Modulate’s technology outweighed potential negative effects. Pappas explained, “This is technology that gives everyone more freedom generally and on net that’s positive.” He recognized, however, “There are ways people use their freedoms to do negative things.” Pappas continued, “By allowing people to access different voice skins we’re giving everyone more freedom in a positive way, but, we are also giving bad actors more freedom to insert themselves in situations more easily where they could do harm.”

For the moment, Pappas and Huffman felt confident that their decision to target video game platforms as their end customers gave them sufficient control over how Modulate voice skins would be used. And yet they were still uncertain about exactly what criteria to apply when working with potential customers. Huffman pointed out, “From Day One, we were aware this technology could be misused in principle and therefore we always needed to think about ethics.” He conceded, “It’s just as we have spoken to more people, they have brought up things that we hadn’t necessarily otherwise considered.” Hence, given the rapidly evolving nature of the voice skin technology they wondered, how could they ensure that the decisions they made today would still have good future outcomes and enable them to be a profitable business?

Exhibit 1: Modulate Website Ethics FAQ

Ethics FAQ

Modulate firmly believes in the social benefit voice skins can provide, but we also understand that they come with risks, and we take seriously our responsibility to ensure our voice skins are not misused. If you have any questions about our approach to these issues, please reach out to us at ethics@modulate.ai.

I don't want anyone using my voice without permission. Will Modulate let people do this?

No, Modulate will not enable voice fraud.

There may be some applications where customers will wish to use Modulate to create voice skins which are inspired by real people. If the customer does so, though, we'll be making sure they have gotten permission to use that voice before we create their voice skins.

Should you discover that someone has misused your voice, and we failed to catch it, please email us at ethics@modulate.ai, and we'll work with you to identify the user responsible, and pursue any necessary punishment and/or remediation.

This sounds scary. Couldn't it be used for fake news?

Don't worry, we've got you covered here too!

Modulate allows you to create audio that sounds real - but that doesn't mean there's no way to know it's fake! We watermark all the audio we generate. While we can't confirm whether an audio clip was synthesized or altered elsewhere, the presence of our watermark makes it easy to identify any content created here, and ensure it isn't treated as evidence of a real event.

I want to learn more about how you're thinking about ethics and your responsibility. Where should I look?

I'd start by checking out our blog post, which discusses our thinking about these problems in a bit more detail. You can also reach out to us directly at ethics@modulate.ai with any questions or concerns.

I appreciate what Modulate's doing to ensure your technology is used responsibly, but others might not share your sense of responsibility. Is there anything we can do more generally?

It's sad but true that, just as Photoshop meant you could never be certain about images, we must now grapple with the fact that we cannot simply trust audio. But just as we still have trustworthy images, we can still have trustworthy audio - it will just require each of us to pay a bit more attention. If you're concerned, the most valuable thing you can do is to have this conversation, so that everyone knows to be more careful with audio, and to take common-sense measures like checking where the audio is coming from, whether or not there are background noises that fit with the situation, etc.

As a consumer, we also recommend you think carefully about using any synthesis technology - whether for audio, video, images, or anything else - that doesn't offer a clear commitment to designing and releasing their technology responsibly. At the end of the day, ethics is hard - we all need to be proactive to shape the world we want!

Source: Company website, https://modulate.ai/ethics, accessed September 2019.

Teaching Note: “Skin in the Game: Modulate AI and Addressing the Legal and Ethical Challenges of Voice Skin Technology”

Summaries

These materials offer a basic background about the challenges raised by the rapid advances of AI and highlight issues that businesses may face as they develop and sell AI products. Readers gain a foundation to understand the inherent tradeoffs in moving from AI principles to the implementation and execution of AI technology, and then evaluate organizations’ approaches to addressing such tradeoffs. And, if an instructor so chooses, readers can be pushed to develop their own set of recommendations to deal with these challenges.

The case, “Skin in the Game: Modulate AI Addresses the Legal and Ethical Challenges of Voice Skin Technology” follows the entrepreneurs Carter Huffman and Mike Pappas as they seek to bring their technology to market. By innovatively applying concepts from artificial intelligence systems called Generative Adversarial Networks (GANs) the two men created a novel approach for transforming one voice to sound like another in real time. Huffman and Pappas recognize that their new technology might be also used maliciously yet they are a tiny start-up with competing priorities and limited human and financial resources. The case asks readers to consider not only how Modulate can push forward and successfully meet the typical demands of a technology start-up—technological excellence, preserve first to market position, satisfying investors—but also uphold a code of ethics.

The background primer covers some of the related legal and social issues raised by the constant advances in AI systems that can process more data, more quickly, at less cost and with less human intervention than ever before. Advanced AI systems capabilities have outpaced human abilities in various areas (e.g., discerning patterns recognizable only to machines). In an effort to tackle these AI issues, a range of new frameworks has been proposed by companies, governments, international organizations, and non-governmental organizations. Notably, however, many of these principles and frameworks have been high-level and abstract. This leaves unresolved important questions about how these principles can be implemented at practical level. How can these actors operationalize those principles within the context of the fast-paced, results-oriented competitive marketplace? What issues arise in organizations around AI? How have these AI issues been resolved?

Positioning

These materials are for those who want to understand the complex challenges of developing, deploying, and using AI technologies in an ethical, responsible way, within a highly competitive commercial environment.

Potential audiences include:

Companies can use this case in-house to help both employees and executives gain greater awareness of the tradeoffs and challenges inherent in the use of AI.

Startups can use this case to help them find mentors and investors who can support them in developing their technologies responsibly.

Instructors with a focus on law, business, technology and/or ethics can use this case with students at all levels (graduate, college and even high school) to illustrate the challenges of translating principle into practice in the real world.

Learning Objectives

Case readers will be able to identify and consider:

1. Both intended and unintended potential impacts from the development, adoption, and use of AI technologies.
2. Key stakeholders involved in the development and deployment of AI systems, and describe how and why the interests of those stakeholders might diverge.
3. The possible conflicting interests within organizations between maximizing their business objectives (such as revenue growth, market share, new users, etc.) while also adhering to and advancing AI ethical principles.
4. Approaches organizations can take to help balance and address competing priorities between business growth and ethical issues.

Pre-Class Assignment

Read the case, "Skin in the Game: Modulate AI Addresses the Legal and Ethical Challenges of Voice Skin Technology" as well as the primer.

Consider the questions below as you read the materials.

Reflection Questions

1. Who, if anyone, has responsibility for how customers use a company's technology? The company? Investors? Entrepreneurs? Explain.
2. Imagine that in a year, Modulate's competitors quickly catch up with their technologies. How might such market pressures impact Modulate's decisions and principles? Is the ability to do the "right thing" a luxury only afforded to the market leader?
3. At what point does a technology cease to be controlled by the creator and become controlled by the user? Is it reasonable for Modulate to determine how and in what ways their customers (and the end users) can use their product? Feasible? What might be the ramifications of this approach?

Summary Teaching Plan (90 minutes total)

Discussion 1: Access to Mentors (20 minutes)

The case touches on the influence that mentors can have on a young company. As they began to think about how to pitch their nascent business, Pappas and Huffman turned to MIT's Venture Mentoring Service (VMS) for advice, a service for MIT students and alumni that connects startups with more experienced entrepreneurs and business leaders. Through their interactions with VMS the men realized that having a novel technology was not sufficient for having a viable business. Working with the VMS mentors helped them understand the need to identify a specific market and business model for their voice skin technology. Importantly, through these mentoring conversations, Huffman and Pappas were able to identify business models that could both be successful but also limit the potential misuse of their technology. VMS' guidance was especially critical when Modulate sought venture capital funding. Access to this mentorship helped them identify markets and business strategies that would both be attractive to investors and achieve their goals for responsible business practices. Different mentorship (or no mentorship at all) might have made it harder to find markets or a business strategy that balanced these factors.

For Discussion:

1. How should startup founders identify and select mentors? What criteria should be factors in selecting mentors?
2. For founders who do not have access to something like VMS, what resources can they use to get mentorship?
3. As Modulate grows, what approaches could they try to continue to solicit and receive objective, external feedback and mentorship about the ethical implications of their work?
4. How should founders communicate their priorities to mentors? Imagine that you're a founder whose mentor has just shared that they disagree with your commitment to preventing misuse of your technology and advises you to focus exclusively on revenue generation; how do you go about evaluating that advice?

Discussion 2: Role of Investors (30 minutes)

Many investors in early stage companies seek a quick return on their investment—usually when the startup is acquired or when they IPO. For a small company like Modulate, every dollar and minute spent preventing the misuse of their technology, is money and time taken away from improving their technology, developing new features, or acquiring customers. Given that Modulate did not have any customers at the time, potential investors could have demanded that Modulate prioritize developing their technology and attracting customers before worrying about any hypothetical misuse. Addressing hypothetical misuse of technology may be a good long-term investment, but is often not aligned with venture capital's desire for a quick, high-value exit.

However, partially through luck, Pappas and Huffman found venture capital investors who valued their approach to addressing ethical concerns. These investors—unlike some others Modulate spoke with—regarded Modulate's proactive attempt to prevent malicious use of its technology as a differentiator that gave it a competitive edge over other companies in similar spaces.

For Discussion:

1. Should venture capital investors and funders take AI ethics into account when evaluating companies? Why? Why not?
2. Describe possible power imbalances between startups and investors? How might this power differential impact decisions about product design and business strategy?
3. Who, if anyone, has responsibility for how customers use a company's technology? The company? Investors? Entrepreneurs? Explain.
4. How might start-ups find investors who are aligned with their overall objectives, including ethical objectives? How critical is this alignment?

Discussion 3: The Costs of Ethics (15 min)

At several points, Huffman and Pappas make choices about the development of their technology and their business model that may—at least in the short term—reduce their potential market and ability to maximize profit. Significantly, they chose not to make a direct-to-consumer downloadable app even though many encouraged them to do so, arguing it was a quick and easy revenue source. Instead, the men chose to pursue the video game market where they felt there was greater opportunity to create ethical guard rails around user behavior. Additionally, they invested time and energy in making a watermarking system, did not allow customers or users to make their own voices, and evaluated the appropriateness of new voices on a case-by-case basis. They recognized however, and even expressed some ambivalence, that their desire to make ethical choices might create technical and market disadvantages on which their competitors might capitalize.

For Discussion

1. Imagine that in a year, Modulate's competitors quickly catch up with their own technologies. How might such market pressures impact Modulate's decisions and principles? Is the ability to do the "right thing" a luxury afforded only to the market leader?
2. Imagine that you're the CEO of a Modulate competitor deciding on your company's business strategy. Your company is newer, smaller, and less well-known, but the potential market is huge and no voice skin company is very large. Why might you want to focus on maximizing growth even if it means potential misuse of your product? Why might you want to focus on being even more focused on ethics than Modulate, even if it means slower growth or a smaller market?

Discussion 4: Control of Technology Use (10 min)

The case describes the ways Modulate sought to maintain maximum control by limiting how customers could use their technology. First, they limited their target market. By concentrating on a single market (video games) they narrowed the scope of possible use (and misuse) of their product. Second, they focused on business-to-business sales, rather than direct to consumer sales, which gave them greater control over the use of their technology. Third, when questions arose about their technology's use they addressed them in all company meetings; everyone in the (small) company had the opportunity to voice possible responses. While not scalable, this approach had advantages; challenges could be addressed and resolved in real time with buy-in and support from team members.

For Discussion

1. What challenges will Modulate face with its current approach? How sustainable is its current approach? In the long-term, what approaches could Modulate consider? When is it important for a company to include every employee in decision-making, and when is that not important?

2. When does a technology cease to be controlled by the creator and become controlled by the user? Is it reasonable for Modulate to determine how and in what ways their customers (and the end users) can use their product? Feasible? What might be the ramifications of this approach?

3. Imagine that instead of limiting uses of their technology, Modulate took a more hands-off approach, allowing users to decide for themselves how the technology should be used. Some people might argue that Facebook has taken this kind of more hands-off approach with regards to policing user content on their platform. What might be the financial, market, and social ramifications of Modulate taking a less restrictive approach?

Discussion 5: Conclusions (10 min)

This case presents an opportunity to consider the complexities and challenges of bridging principle to practice. Even well-intentioned founders who run a small company and control almost every aspect of their business and technology face significant hurdles—economic, organizational, operational, and competitive—that make it challenging to develop and deploy AI technologies responsibly. The immense pressure to move fast, be innovative, and maximize profits, make it sometimes seem as though addressing the complex ethical challenges inherent in their technologies is a luxury of time, money, and mindshare that they simply cannot afford. This case, however, demonstrates ways that these hurdles might be overcome, and in so doing, an organization may actually be better positioned in the long run for having made investments in technology, operations, and people, as well as by having mission alignment among its investors, employees, and even customers.

Final Takeaways

- The decisions that companies make are shaped not only by the founders and executives, but also by the investors, customers, mentors, and even competitors. For that reason, a decision to prioritize responsible use of technology, or deemphasize it, is often the result of numerous, complex, interlocking, and even competing factors; it is often an oversimplification to pin such decisions on "good" or "bad" leaders. By the same token, changing company behaviors is more than changing a single factor—it often requires systemic changes to venture capital, valuation, training, mentorship, and more.

- A company's approach to AI ethical challenges may change over time. The factors described above that influence a company's decision are not static. Founders leave or change their minds. Funding and funders change. Customers' needs and desires change. Competitors emerge, succeed, and fail. All of these changes will influence a company's approach to these challenges.

- For many companies, particularly smaller ones, investing money to address AI ethics is a zero sum game. Every dollar spent tackling an ethical challenge is a dollar not spent and a loss for new features and services. This is not to say that investing in AI ethics is a bad decision, but only that it creates difficult questions for the founders, executives, and investors.

Background Primer: “Skin in the Game: Modulate AI and Addressing the Legal and Ethical Challenges of Voice Skin Technology”

A Companion to the Case Study

I. Introduction

This research memo is a companion to the case study, *Skin in the Game: Modulate AI and Addressing the Legal and Ethical Challenges of Voice Skin Technology*, providing a brief background about technical, social, and legal elements raised by the case. This research memo does not provide an exhaustive analysis of these topics; instead it provides a background for those who wish to explore the challenges companies such as Modulate face when developing, deploying, and using AI technologies.

Many of the most challenging issues that organizations face when creating or using emerging technologies have no clear solutions – this is what makes these questions challenging! Likewise, this memo does not offer legal advice or provide an ethical roadmap. This memo provides a helpful technical, social, and legal background while the case study offers a context in which to engage and explore these truly unresolved and challenging questions. This memo offers readers a way to grapple with the difficult questions raised by the case study with less of a chance of getting lost in technical or legal questions that experts have already addressed (e.g., “who owns the copyright in a synthetic voice?”).

II. Technical background

Generative Adversarial Networks (GANs) and digital media: A Primer

At the broadest level, artificial intelligence (AI) encompasses many different technologies that allow machines to carry out tasks that seemingly require human cognition and reasoning.1 In 2020, when people discuss AI, they are often referring to machine learning (ML). Machine learning is a set of techniques that use algorithms and statistical analysis to analyze data to create models and “learn” from correlations.2 One form of machine learning is a neural network – a set of mathematical values that model the human brain’s network of neurons. Another more recently developed approach uses two competing neural networks in a generative adversarial network, or GAN. In a GAN, one network’s goal is to generate the most believable “fakes” (fake audio, fake images, etc.), while the other network tries to identify the fake data.

[1] http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html

[2] https://deepai.org/machine-learning-glossary-and-terms/machine-learning

Through this process of competition and iterative improvement, the GAN learns how to create synthetic data that is difficult to distinguish from the original data. Practically speaking, a GAN can not only learn to create believable images and sounds of anything—human speech, human faces in photos and videos, and more—but also learn to alter them in realistic ways.

Technology critics warn that without appropriate safeguards, the proliferation of GAN-generated media could mislead people and drastically erode their trust in digital media. Paradoxically, the very quality that gives GANs their great potential is also what makes it so hard to develop technical solutions against them; a successful GAN learns how to fool automatic detection. At present, there is no simple existing technical solution to detect and deter against synthetic media generated with GANs. In order to work towards a future where we balance GAN's risks with its benefits, more concrete and holistic solutions that respond to the crux of the issues surrounding such technologies[3] must be developed.

III. Societal issues and context

Machine learning and AI advances have introduced technologies that increasingly complicate the relationship between the "real" and "imaginary." As with other emerging technologies, AI-enabled digital media technologies—which encompass everything from photo altering software (e.g., Photoshop), computer graphics, deep fake videos, synthetic audio generation software, voice alteration software, and much more—present both opportunities and risks for new forms of human expression.

The development of digital technologies and the expression of human identities have always been closely linked. Digital photo editing software, for example, allows people to create alternate appearances and realities that are seemingly "real." Photoshop, as the first software to bring photo editing to the masses, was met with explosive enthusiasm, but also concern about its impact on our relationship with digital media and truth.[4] Digital technologies make the suspension of disbelief easier, blurring the line between the realms of the "real" and "imaginary."

Life-like digital creations can have a powerful liberatory quality, freeing people from the bounds of their physical constraints. Female gamers, often targets of harassment in online gaming spaces, can use voice modulators to cloak their real-life gender as a masking strategy. Similarly, some have argued that online spaces can empower transgender users by enabling them to "take off previous identities in favor of chosen identities that reflect their claimed personalities."[5] This freedom could enable people to challenge long-held assumptions about identity, and even help advance a cultural understanding of identity markers such as the social construction of gender.[6] Similarly, digital technologies' potential for such "identity play"[7] could allow for increased understanding of, and compassion for the challenges that people with other identities face as they navigate the world. However, scholars such as Lisa Nakamura, Lori Kendall, Kishonna Gray, and others have more recently drawn attention to how these technologies are created and embedded within offline societal structures and norms, potentially limiting their transformative effects.[8]

Ultimately, the deployment of these technologies alongside other digital technologies will complicate our understanding about identity, truth, and reality. Identity-changing digital technologies present incredible potential for misuse that could erode trust in digital media. AI and machine learning have

[3] https://spectrum.ieee.org/tech-talk/robotics/artificial-intelligence/will-deepfakes-detection-be-ready-for-2020

[4] https://www.huffpost.com/entry/the-years-in-retrospect-how-photoshop-has-shaped-the-world-of-graphic-design_b_11848126

[5] *Living the VirtuReal: Negotiating Transgender Identity in Cyberspace*

[6] *Bent Gender*, http://www.cios.org/EJCPUBLIC/005/4/00545.HTML

[7] *Cyberspace and Identity*

[8] *Cyberspace: The Performance of Gener, Class, and Race Online*. Lori Kendall

been used to generate convincing fake audio, images, and video (called "deepfakes") using recordings from public figures. For example deepfakes have been created with audio from Joe Rogan's podcast[9] and from video of President Obama[10]. These fakes can increasingly be created in real-time.[11] These are just a few well-known examples, but a simple web search yields thousands of similar results, and many of these have been used to generate sensational 'fake news' items that spread virally through social media.

Convincing deepfakes signify a great technical achievement, but also significant disruption to societal norms, both negatively and positively. For example, augmented reality applications that use facial recognition to put real-time video 'skins' and masks on people may be amusing, but the software (and the generated data set of users that willingly contributed their faces) may also be used to train better facial recognition algorithms for surveillance. Or in another example, voice modulators and synthetic audio generators can help female gamers mitigate online harassment by hiding their gender, but the same technology can be used to more effectively impersonate people for fraud or phishing scams.

New technologies are never developed in a vacuum. They must be thoughtfully developed and deployed with an understanding of this context, in order to minimize their harm.

IV. Legal issues

AI technologies, particularly when used to create synthetic voices, images, or videos can provoke a range of legal questions. "Audio skin" technology—like the other technologies discussed above—provides both opportunities for positive and negative uses. The technology offers us the freedom to choose different voices and explore different identities, but depending on intent, the outcome of these choices could be very different. Nefarious uses of this technology pose several legal and ethical issues.

Intellectual Property

Intellectual property law grants exclusive rights and protections to the owners of intellectual property. These rights generally apply to creative works, but can also apply to aspects of a person's identity (voice, visual likeness, etc.), or product brands. Intellectual property law tries to balance the interest of content creators with the interest of the public in accessing and using those works. Intellectual property law includes several different areas such as copyright and trademark.

Copyright

In order for something to have copyright protection, it must be an "original work of authorship" (e.g., literary works, musical works, graphic works, sound recordings, etc.) and it must be fixed in a "tangible medium of expression" (e.g., a literary work that is set down in writing in a book, or images or sounds that are recorded digitally or on film). Generally, spoken voices in and of themselves, are not copyrightable because sounds alone are not fixed in tangible medium; only particular recordings of sounds are subject to copyright. But a voice that is pre-recorded, like a clip of a famous actor's voice copied from a movie, would be copyright protected. Hence a technology that offers users the ability to adopt a celebrity's voice or allows them to record others' voices and adopt them as their own, could run afoul of copyright.

That said, something that might otherwise be a copyright violation may be allowed if it is considered

[9] https://www.theverge.com/2019/5/17/18629024/joe-rogan-ai-fake-voice-clone-deepfake-dessa
[10] https://www.buzzfeednews.com/article/davidmack/obama-fake-news-jordan-peele-psa-video-buzzfeed
[11] https://www.engadget.com/2019/05/15/google-translatotron-direct-speech-translation/

"fair use." There is no bright line rule for what constitutes fair use. Instead courts balance several different factors on a case-by-case basis to determine whether a use is fair. For example, if the purpose of the use is commercial, it will weigh against fair use, but a use for non-profit educational purposes will incline towards fair use. Fair use can include things like criticism, comment, news reporting, teaching (e.g., using copies of a work in a classroom), scholarship and research. For voice-skin technology, that means that a school designed platform that teaches empathy by letting students' use different voices might have a stronger fair use case than a purely commercial use of voice-skin technology. Another factor in a fair use analysis would be "transformative use" – whether the copy transforms the original by adding some new expression or meaning. An audio skin technology that just makes users sound exactly like celebrities, would probably not be sufficiently transformative.

Trademark & Right of Publicity

Particular voices may have trademark protection. Trademarks are words, symbols, phrases, or sounds that identify a particular manufacturer or seller's products or services. Trademarks protect a brand in order to protect consumers from deception or confusion and to protect producers' property. Courts have ruled that distinctive and unique voices may be protected by trademark law, because a person's voice can be a distinctive indicator of their identity. Voice skin technology that either copies or imitates certain distinctive, well-known voices could be abused to create false endorsements that would run afoul of trademark law.

Similarly, the right of publicity allows each person to control the commercial use of their identity. It prevents others from using aspects of one's identity, such as one's voice, in a way that may damage its commercial value. For example, a voice skin technology developer that uses a celebrity voice to draw attention to their product can constitute an infringement. Right of publicity may even extend to non-celebrities. For example, using someone's voice skin in advertisements shown to their friends may violate their right to publicity.

Secondary Liability

A voice skin technology creator may also face legal issues if their users misuse their products to violate others' intellectual property. This kind of liability is called "intermediary liability" because the technology or platform creator serves as an enabler, conduit, host, or forum for someone else's legal violations. There are two kinds of intermediary liability. A technology provider may be contributorily liable if they (1) had knowledge of the infringement and (2) meaningfully contributed to the infringement, and may be vicariously liable if they (1) had the right and ability to control or supervise the infringer's acts and (2) directly financially benefited from the infringement.

A technology producer can be contributorily liable when they make available facilities or tools with which users can infringe copyright. In the defining case addressing this issue,[12] television show producers sued Sony for contributory copyright infringement for manufacturing video tape recorders (VTR) that allowed users to record programs for viewing at a later time. The Supreme Court said that the sale or manufacture of such tools does not constitute contributory infringement if the product is capable of substantial noninfringing uses – that is, if it can be widely used for legitimate and unobjectionable purposes. Subsequent cases have held that manufacturers of tools or technology can still be contributorily liable when there are legitimate uses for their products, if there is evidence that the manufacturer intended the product to be used to infringe copyright. Intent to facilitate infringement can be shown, for example, by advertising or instructing users on unlawful uses of its services or

[12] *Sony Corp. of America v. Universal City Studios, Inc.*, *464* U.S. 417 (1984).

tools. In determining whether there was such intent, courts will also consider evidence of whether the platform attempted to profit off of illegal uses and whether it made any attempts to stop infringing activity.

Other Possible Legal Issues

Creators of voice skin technologies may also face legal issues related to creating dangerous conditions for others. For example, they may face product-liability or negligent supervision claims. Product liability claims are usually made against manufacturers of defective goods that lead to injury. If someone is harmed because the technology developer failed to exercise reasonable care, the manufacturer can be liable for damages.

Negligent supervision claims usually apply to those who provide physical spaces for people, like shopping malls. The owners of those spaces have a duty to take reasonable care to safeguard visitors from physical harm. For instance, if there are reasonable, cost-effective, and not unduly burdensome steps the owner could have taken to prevent a reasonably foreseeable crime, and the owner did not take those steps, the owner could be liable for negligent supervision. Although it is uncertain if such claims would prevail in court, it is possible to imagine voice skin technology being used to commit fraud, and the victims of such crimes bringing claims against the developers of the technology.