“The term ‘programmer’ also used to refer to a human, someone writing programs the old way, with keyboard and mouse. Now the craft of computer programming has been automated thanks to voice-operated programming (VOP).”
EMMA BEAUXIS-AUSSALET

From Dark Roots to Shared Routes

We all struggle with computers. If only we could talk to them, so that they could understand what we want without having to reprogram them. Controlling computers with our voice sounds much better, unless computers control us with their own customized voice.

INTRODUCTION

The bittersweet utopian text “From Dark Roots to Shared Routes” explores a future where Natural Language Processing (NLP) technologies, formerly used to manipulate people with commercial and political campaigns are being repurposed for greater good. To achieve this transformation, courageous civilians and politicians took action. Nefarious uses of NLP technologies were banned but not the technologies themselves. Better uses of the technologies lead to voice-operated programming: a technology that enables anyone to program their own computers by simply talking to them. Yet such voice technologies have side effects that threaten our agency and the fabric of our society. NLP technologies, social media analysis, and microtargeting can seriously damage our autonomy and our democracy. Microtargeting adapts...
marketing messages to specific individuals and trigger specific personality by prompting tailored psychological cues. This nudging affects our autonomy, as our decisions are based not only on information but also on emotions.

Voice technologies can make social engineering even more efficient by adding nonverbal psychological cues. Social media analysis can make social engineering more pervasive by identifying the language patterns in the social circles of the targeted individuals. Generative adversarial networks (GAN) can then be used to generate voice messages that sound just like the specific individuals in targeted social circles. Then, with chatbots, humans could engage into mind-bending microtargeted conversations, without even realizing that they are talking to a computer. Yet, as this essay shows, the same technologies can be used to enable humans to better control their computers rather than the other way around. This story presents some of the social impacts these AI technologies already have, and may have in the future. It explores the steps we might take to progress from tacit use of AI for evil, to public scrutiny of computing systems, and to open use of AI for the greater good.
A long time ago, the term “computer” used to refer to a human, someone doing computations the old way, with paper and pen. The term “programmer” also used to refer to a human, someone writing programs the old way, with keyboard and mouse. Now the craft of computer programming has been automated thanks to voice-operated programming (VOP). We can carry out almost any task by simply talking to our computers. Our voice operators find and execute the programs we need and write the programs that are missing. We barely need to move a finger anymore, for a click or a keystroke.

Operating computers with our voice is much healthier: no more slouching and broken backs, but improved breathing and blood oxygen balance. And operating computers is also more accessible than ever, as voice operators can rapidly adapt to understand many accents and dialects. An incredible diversity of people, including children, can design and operate their own information systems. VOP has been deemed “the most empowering technology of the century” that has “greatly improved the autonomy and efficiency of millions of workers” and “unleashed the creativity of human kind” (OECD, Development Co-operation Report, 2039).

However, few of us remember the dark side of VOP history. It stemmed from the evils of surveillance capitalism, from technologies that compromised our democracy. Let’s explore the difficult truths behind the genesis of VOP and reflect on the ethical and political issues of its past ... and present.

**GENESIS: TO HOAX AND TO COAX**

Twenty years ago, deep fake technologies were plaguing the internet. They were used to forge videos of politicians and manipulate elections, but also to tweak advertisements to target our specific psychological profiles. Neuro-marketing companies could generate highly personalized ads, videos, and even newspaper articles without any human actually writing or shooting them. Their fabricated texts and videos had a cunning realism and a strong power of influence.

For political campaigns, the fakes targeted those who could vote for political opponents. They were designed to break the voter’s trust and the politician’s decorum. The most insidious fakes staged embarrassing incidents, with politicians in ridiculous or demeaning situations. The fakes could spread rumors indefinitely as no attempt to debunk them was ever credible—or credited—enough to quench our thirst for mockery and controversy. Social media platforms embraced fake news in their business
models, giving visibility and advertisement revenue to all parties, and
dodging any editorial responsibility.

In 2024, the US presidential election was gravely biased by a dis-
information campaign. A seemingly real video of President Bernie Sanders
was released just 72 hours before the first polling stations opened. This
extremely well-crafted video, accompanied by the release of hundreds
of fake documents and newspaper articles, led the American public to
believe in a conspiracy with the Chinese government. The attack framed
Democrats receiving bribes to implement their newly adopted Green New
Deal with Chinese green technologies. It dramatically swayed the election.
Polls were predicting Bernie Sanders’ reelection by a large majority, but
instead Stephen Baldwin received the majority of votes.

It took weeks to fully debunk this disinformation campaign. It was based
on an intricate network of fake documents, and their dissemination had
started months before the election. The fakes manufactured many counter-
claims and misleading evidence. By the time the truth emerged, the US was
in the midst of a dire constitutional crisis. Most civilians and politicians
were in favor of annulling the election and organizing another one, but
no legal framework could allow it. The crisis ended with the withdrawal
of Baldwin, who terminated his short-lived political career to facilitate the
investigations and protect his reputation and assets. Four of his campaign
managers were later convicted of conspiracy and defamation.

This deep fake scandal was a historical crisis that outraged citizens all
over the world. The technologies at play drew the attention of lawyers, AI
experts, policy makers, scholars, hackers, and journalists. In an unusual
synergy of multiple disciplines, they carefully scrutinized other uses of
deep fakes. They soon uncovered another deceitful technology: social
language modelling.

This technology can adapt the tone of texts to mimic the language of a
person’s social circle. It can produce marketing or political messages that
sound like our friends or family members. Who could resist buying clothes
that are recommended by someone just like your partner? Or voting for
politicians endorsed by someone just like your best friend? The discovery
of social language modelling (SLM) was a major scandal. Citizens were
appalled by the practice but no legal framework could ban it until the 2030s.

All SLM retailers were trafficking the exact same technology, provided
by a group of hackers who remain unidentified to this day. They operated
through a shell company named Sheepshape. They framed their services
as conventional AI using no private data, but applied extremely strict non-disclosure agreements. The use of AI had to remain confidential, and Sheepshape services were officially basic data science. Behind such extreme trade secret protection lay spectacular privacy violations. In practice, the language models were trained with highly personal data such as private emails and chats. With these, Sheepshape could produce extremely efficient communication campaigns, far surpassing conventional neuro-marketing.

Sheepshape had backdoors for accessing private communications on at least Gmail, Facebook, and Twitter. It relied on a handful of corrupt employees at major IT companies. They disseminated spyware at the lowest levels of cloud platforms: in the compilers of data backup and encryption components. This complex technical scheme was uncovered by a team of investigative journalists, hackers, and researchers. In a remarkable retro-engineering feat, they analyzed how marketing messages were tailored to real and simulated individuals, and identified which personal communications were used to build the language models.

In the wake of the Sheepshape scandal, an anonymous source leaked the SLM technology to the public in 2026. The leak enabled SLM to spread into most marketing campaigns, but this time the language models were trained using public data from social media. On the bright side, the leak also enabled great innovations that improved voice interfaces. It also prompted the adoption of crucial regulations. In 2031, the United Nations Security Council announced a ban on blacklisted AI for public influence purposes on the grounds that such manipulative AI systems threaten human dignity and autonomy. It was clear that companies that owned public influence technology had gained too much power over our democracy and economy.

The blacklisted technologies included SLM, deep fake, and recommendation systems designed to maintain and polarize audiences within filter bubbles. The models and algorithms in question were published and documented to enable public scrutiny and international cooperation. Hackers and academics soon repurposed these back-alley monsters of neuro-marketing and gave birth to the first talking computers. Back then, we did not imagine the coming revolution, just as we did not imagine the internet revolution in the twentieth century.

FROM SOUND TO SOUND: THE RISE OF TALKING COMPUTERS

Talking machines are an old human fantasy. For ancient Greeks, the god Hephaestus crafted automatons that mimicked humans or animals. Interestingly, he also crafted Pandora and her box. The first talking
machines appeared thousands of years after those ancient myths. And they did not live up to our expectations. They made countless errors and misunderstood many populations with accents that did not comply with some arbitrary standards.

Unexpectedly, deep fake and SLM technologies were the missing pieces. They made talking computers able to understand each user's way of talking, with different pronunciations and vocabularies. Talking computers became highly personalized, almost perfectly tailored to their individual users. Their errors became rare and easy to correct. We grew more and more comfortable talking with computers; it became like talking with familiar colleagues.

Voice operators, as they are formally called, have completely transformed our ways of living. Voice-ops, as we usually call them, handle much more than computations. More than the world of data and programs, they now handle a large part of our intellectual world. They read and write our books as well as our computer code. They are our curators, our secretaries, and at times, our companions. Today, in 2041, we spend most of our day wearing noise cancelling headphones, at work or at home, seamlessly talking with humans or computers, under the spell of omnipresent vocal presences.

Human communication has taken many shapes and forms through history. Thus far, as we now turn our relationships, our books, and our programs into vocal presences, our communications have evolved from sound to sound. From prehistoric to digital humans, information has mutated from music to speech, to writing, to computing, and eventually, to the realm of sounds again.

In the early 2030s, deep fake and SLM were first used to make toy apps for entertainment, but the field quickly professionalized to develop voice operators for customer services. Having voice operators was a sign of modernity, and most companies implemented them to keep up with the trend. To foster the success of voice operators, the software industry adopted standards for typical voice commands to trigger menus and buttons, or to fill in forms. Thanks to these standards, most websites became voice-operable by the mid 2030s.

Developers started to enhance their own programming tools with voice interfaces. Using the voice turned out to be a great help when developing computer programs. Besides relieving RSI symptoms, voice interfaces allow for more creative thinking. As the body is unconstrained by keyboard and mouse, the mind can focus better on the high-level design of algorithms.
Developers around the world kept collaborating on open-source voice interfaces using state-of-the-art AI, such as deep fake and SLM technologies to handle the different jargons of programmers.

Academics coined the term voice-operated programming (VOP) in 2034 and developed intelligent dialog protocols to translate high-level descriptions of our goals into low-level computer operations. The dialog protocols tell VOP agents how to lead a conversation until enough information has been gathered to specify the programs we need. VOP agents would then orchestrate and execute those programs. Should a program be missing, VOP agents would write the code, deploy the program, and do tests and debugging.

At first, communicating with VOP agents was rather unnatural and required an obnoxious computer vocabulary. For instance, let's say we owned a business, and we asked our voice-op: “Send a reminder about their bills to the clients who are late”. The voice-op may ask: “What do you mean by ‘send a reminder’?”. We would then need to say something awkward like: “The reminder is a mail. It sends the unpaid bills to a client. Apply a summarizer to the bills. Set the summarizer parameters with result length to 3 sentences and word cap to 100. Use the result of the summarizer as the text content of the mail.”

The VOP results may have failed. It could have written a reminder like: “Dear client, you have 3 unpaid bills hereby attached. Total due is $774.37 (incl. $82.07 tax and $45.40 shipping). It represents 5.17% of our net revenue for Week 51 of 2035.” That may be the correct summary for a business owner, but not one to send to clients. To improve the summary, we would have needed to program another summarizer via lengthy dialog with our voice-op.

Today our dialog is more natural and more personal. We simply say “Drop these clients a note on their bills.” VOP has achieved this level of technological development thanks to well-established standards and design guidelines. With good compatibility standards, VOP systems easily combined a variety of languages. Models of human and computer languages became interchangeable and personalizable. The VOP industry grew from new language models, to new business models, to new success stories.

VOP pervades most of our professional and personal lives. Our daily activities are largely voice-operated. Anyone can easily construct complex AI systems, and most children are fluent in VOP by the age of 12. However, this groundbreaking technological progress is not without side effects.
VOP entails more complex and lengthy computations compared to traditional programming. These extra computations are greatly contributing to the climate crisis. This, too, is part of the dark roots of VOP: its energy consumption and consequent pollution. Let’s explore the challenges we need to tackle for limiting the computing resources and thus the ecological costs of VOP systems.

**PAYING THE BILLS: THE COMPUTATIONAL AND ECOLOGICAL COSTS**

Voice-operated programming requires all sorts of additional computations to work out the programs it needs to execute and discuss them with us. A first layer of computations is needed to match what we say in our own natural language with standard voice commands. A second layer is needed to translate the voice commands into low-level programming operations. These two extra layers often need much more computations than the actual VOP-generated programs we eventually execute. At a global scale, the energy consumption of these additional computations is staggering. It is comparable to the energy consumption of the entire transport industry.

On top of that, the programs written by VOP consume more energy than programs written by humans. VOP often generates complex software architectures that are not optimal. For example, VOP reuses pieces of code that sometimes include computations that have no purpose in the new context. And even when VOP writes new code, its programs need more computations than programs written and optimized by humans. The capacity of humans to optimize computer systems remains far superior to what machines can do to optimize themselves. But to optimize our programs, we need highly trained human experts. And they are becoming rare.

Most VOP applications do not justify investing in the verification and optimization of the programs, thus the job market for human programmers has greatly reduced in the past 5 years. Our programming workforce is now at a minimum, and we are losing our ability to fully control most programs we use. When we program with VOP, we remain quite ignorant of what exact programs we are executing. Our programs can have misconceptions and bugs, on top of suboptimal code and increased computational costs. But there is no one to correct them.
Computing resources are extremely fast and cheap, but they have dire ecological consequences. On a global scale, the carbon footprint of voice operators is enormous. In our damaged ecosystems, plagued by heat waves and severe storms, much concern arises from the ecological costs of our computing systems. All industries are liable for their energy consumption under the UN Climate Control Regulations. Yet we barely measure, not to mention regulate, the energy consumption of the computing industry.

The UN Climate Control Regulations require companies to implement supply chains that consume the least possible material and energy resources. Otherwise, they may be barred from international trade. The criteria and strategies for optimizing resources are a source of constant controversy, of course, but international standards have been developed for most industries. Yet standards for the computing industry are lacking.

The only incentives for the digital industry to reduce its energy consumption are ecological taxes on energy. Since 2026, international agreements have enabled worldwide tax policies similar to value-added tax: a kind of pollution-added tax. But these are not sufficient. We know that taxation policies and the self-regulation of industry fall short of addressing the extent of our overconsumption.

Before the United Nations ratified the Climate Control Regulations in 2038, we long thought that taxing energy and waste would be enough to push industries to adopt more sustainable practices. And indeed, lots of industries improved their practices. But wealthier industries could afford the taxes and practically bought their rights to pollute. Our carbon emissions remained highly unsustainable, public health and entire economies continued to crumble, and natural ecosystems continued to collapse. The public pressure for an effective political response grew into massive worldwide strikes in the mid-2030s.

We then considered a more drastic approach: Technologies with high environmental impact must be restricted. They must be limited to cases where the greater good might justify the ecological costs, for instance, health or safety. The idea was largely approved by the public, and national and international regulations were developed. They ultimately led to the UN Climate Control Regulations. Since then, entire domains have been disrupted, such as the transport industry. But impacting entire domains is actually an advantage: It means that no one is able to use more polluting technologies just because they can afford ecological taxes.

A typical example of these climate control laws pertains to freight. Some resources are shipped across the world on highly polluting cargo boats. The essence of the UN Climate Control Regulations is to ban companies from shipping their resources across the world if local resources could replace
them. This apparently simple principle has tremendous consequences for our economy and our ways of living.

For instance, fruits like bananas are no longer imported to northern parts of the globe, but exceptions remain for restaurants and other cultural purposes. Europeans and Canadians have been ready to sacrifice bananas in their regular diet, and banana producers have been ready to sacrifice their international sales, because these are not sacrifices but necessary transformations of our society.

While we reinvent our economy, opportunities are arising too. Cargo boats have been reassigned to collect plastic trash in the oceans, subsidized by funds from ecological taxes and equipped with innovative technologies for depolluting their exhaust gases. Banana producers export less but for a higher price, they produce less but at a higher quality, and they avoid the ecological damage of high-yield agriculture.

The UN Climate Control Regulations have been implemented gradually, one restriction at a time, one alternative solution at a time, one transformation at a time. The time has come to transform the computing industry, and there are well-known targets for reducing its energy consumption. For example, some advanced AI systems entail a lot more computations but provide results that are only a little better. The essence of the UN Climate Control Regulations would be to ban such AI for nonessential purposes, like advertisement or entertainment, but allow them for essential purposes like medical applications.

The technological choices in the computing industry can be vetted without stifling innovation. On the contrary, it gives a direction for innovation: reducing computational and thus ecological costs. It also pushes companies to modernize their old-fashioned deep learning systems. There is an alternative technology that can greatly reduce the computational and ecological costs of AI and VOP: dip learning—the nemesis of deep learning. Both have more or less equivalent results, but dip learning needs only a fraction of the data and computations that deep learning requires. To understand the disruptive power of dip learning, and how it can make VOP more sustainable, we need to go back to the early days of AI.
OUT OF DEPTH: LEARNING TO LEARN

At the end of the 20th century, the first forms of AI were model driven: They encoded human knowledge into models. The models were built manually and represented the elements of a problem to consider, the rules under which these elements could evolve, and the reactions to adopt under specific circumstances. Such models were often very complex, and very difficult to design and maintain. Many possible corner cases and unexpected situations arose under real-world conditions. And many aspects of the possible situations could not be integrated in models that were already too complex and hard to update.

This type of AI was costly and limited. It promised too much and delivered too little. Investments and business prospects dwindled, and AI became an intellectual curiosity for mathematicians. In the early twenty-first century, another approach became successful: data-driven AI. Instead of using hand-made models, data-driven AI uses samples of data that are manually processed. AI algorithms just have to find ways to mimic the results produced by humans. Algorithms could learn how to mimic the results after repetitive sequences of trials and errors, called training, which basically optimized the parameters of low-level data processing techniques.

Data-driven AI achieved what model-driven AI promised. Almost all our AI systems became data driven. Instead of encoding human knowledge into complex models, we started to rely on simple generic models that were agnostic of any human knowledge. To compensate for the lack of human knowledge, we relied on more data samples and more parameters. With this approach, artificial intelligence became more artificial than intelligent.

Dip learning belongs to both worlds: It is both data driven and model driven. It relies on data samples, but needs much less data than deep learning: It just needs a dip in data. It relies on models too, but not on hand-made models. Dip learning makes its own models: It finds the most logical model from the available data samples.

Dip learning models can be as abstract and meaningless as deep learning models. But they are easier to interpret and correct, manually or automatically. Should new data or manual correction be available, dip learning would refine its models automatically and rapidly. These are the strengths of dip learning: It needs less data and fewer computations, and it provides more adaptive and tractable models.
Another strength of dip learning is that it is designed to deal with incomplete and partial data. Dip learning maintains models of what it knows, but also of what it does not know. It can infer hidden factors that are not directly represented in the data. For example, sales data may show that some clients are often late when paying their bills, but none of the variables in the data may be able to help with identifying late and early payers. Dip learning can infer that it needs extra variables, called latent variables. For example, at least one extra variable would be needed to identify late and early payers, such as net income.

Dip learning can model the interactions between variables and latent variables. In theory, deep learning could achieve similar results with something called hyper-parameter tuning. But it would require innumerable trials and errors, with huge computational and ecological costs.

Dip learning can dramatically reduce the staggering energy consumption of voice-operated programming and AI systems in general. Yet many AI systems have not been upgraded to dip learning and continue to operate with deep learning. Their impacts on our ecosystem are not justified and conflict with the UN Climate Control Regulations. But efforts to regulate the computing industry have faced legal opposition: Companies want to keep their technology confidential, and the law strongly protects their intellectual property.

We have achieved tremendous successes in our quest to control our impact on the planet, and remodel our industry and economy. We also achieved tremendous successes in developing breakthrough technologies for voice operators. Through international cooperation, we have revolutionized our technology and our economy. We have such a great potential to extend the cooperation on technologies and regulations. Couldn't we go one step further and agree on restrictions of AI technologies that overconsume computing power? Couldn't we agree on methods to verify whether restricted AI technologies are used in a computing system, while leaving the exact AI models and system architecture confidential?

Even if we addressed the ecological problem of VOP with dip learning and complied with the UN Climate Control Regulations, we would not be done with addressing all problems with VOP. We must also beware of how VOP impacts our culture and our relationships. Let's reflect on how VOP is transforming our society.
WHAT REMAINS UNHEARD: THE HUMAN COSTS

Voice operators are transforming our society because they remodel our work environment, our access to information, and ultimately, our sense of community. At work, voice-ops have introduced great acoustic stress. Offices are like a permanent meeting. Whether colleagues interact among themselves or with their computers, work sounds like a gigantic call center.

To muffle this sound jungle, we are constantly equipped with noise-cancelling headphones. And we are thus constantly reachable for a call from our colleagues or our computers. We stay tuned into the sound flow of humans and machines. Unable to escape the silence haven of our noise-cancelling headphones, we surrender to the acoustic tyranny of remaining on standby for our colleagues and computers.

The toll of voice-ops on our acoustic comfort is not the only worry. With the sonification of information, we barely read or write anymore. But when voice-ops read a text for us, we cannot explore its content at our own pace. Voice-ops disrupt the inner rhythm of our reflections. They can render pauses in speech rather naturally and at a controllable pace. Yet the pauses are imposed, and we understand less of a text when hearing it rather than reading it.

On top of this, a lot of our texts are now written by machines: by web searchers, by summarizers, by data analyzers. They design our news briefs, write our reports, and tweak our ads. They write our books too: novels with plots and well-crafted suspense, handbooks with virtual reality add-ons, school books with personalized exercises. Are we losing our ability to write our own literature? Just as we are losing our ability to write our computer programs?

The machines that read and write our texts have a seemingly real personality of their own. They are sometimes awkward but they sound all the more alive, with an awkwardness of their own as part of their personality. Voice-ops can create the illusion of a companion, a seemingly real interlocutor. They can adapt their tone of voice to influence our emotions, for example, to soothe anger and anxiety or to help us fall asleep. They can also comfort the lonely and the depressed.

Perhaps a lot of us find comfort in the predictable behaviors of machines, in their routines, and in their fully personalized character. After all, we let voice-ops handle many social tasks, such as customer services, therapy, and education from kindergarten to university. As machines blend into our
social fabric, we might end up mimicking them as much as they mimic us. Are we talking more like machines? Do we perform computer-like routines with humans too? It would wash out a lot of our human character, a lot of what cannot be reproduced by machines. It would change our culture, our intellect, and our emotions.

We now spend more time talking with voice-ops than with humans, colleagues, friends, or family. And we rarely meet each other face-to-face. We often prefer phone calls, to hear each other with the comfortable sound of high quality headphones, and the familiar peace of noise cancelling. But we are slowly turning human touch into vocal presence.

We are losing touch with what is actually behind the sound flow of humans and machines. We are losing touch with actual humans and actual machines. Do we really know exactly what our programs are doing, and how much energy they really require? Do we really know exactly how our relatives are feeling, and whether they would rather talk to a comforting voice-op than to us?

**EPILOGUE: THE COURAGE OF HOPE**

Deep fake and social language modelling gave machines the ability to mimic humans. Companies and politicians used them to manipulate our economy and our democracy. Journalists and hackers uncovered their misuses. Regulators banned them for intentions of public influence. Developers redesigned them to build powerful voice interfaces, and voice-operated programming gave them a greater purpose.

We overcame the threat of deep fakes and social language modelling to our economy and democracy. Now, we face the consequences of voice-operated programming for our ecosystem and our social fabric. To face the challenges ahead of us, we must have the courage of our hopes. As we hope to restore a livable planet, and develop the sustainable AI systems it requires, we need the political courage to debate and regulate our technological choices, starting with voice-operated programming.

As we hope to undergo tedious negotiations across states and stakeholders to agree on AI standards and regulations, we need the courage to get past the smoke screen of technological complexity. To get past business-as-usual, trade secrets, and economic pressure. And as we hope to succeed in negotiating rules for a sustainable AI economy, we will need the courage to understand each other’s struggles with the dramatic changes that are bound to happen. We will need the courage to get past the comfort of our machine-like routines. To get past the bubble of our headphones,
and their hypnotic sound flow. We will need the courage to deal with each other’s grievances and hopes with more real human connections than currently exists in our muffled voice-operated world, which silences a lot of our humanity.

**EMMA BEAUXIS-AUSSALET**

AMSTERDAM UNIVERSITY OF APPLIED SCIENCE

Emma Beauxis-Aussalet is a senior track associate at the Digital Society School of Amsterdam University of Applied Science, where she works on applying artificial intelligence (AI) to achieve social good. She also focuses on enabling transparent and accountable AI, where bias and error are explicit and understandable for the general public. Her research at the National Research Institute for Mathematics and Computer Science (CWI) developed simplified visualizations of classification errors and statistical methods to detail error and bias in classification data.