*FOCUS ON:*

# THE RISE OF AI

Welcome to the promising — and challenging — era of smart machines

In March 2022, after the large language model (LLM) ChatGPT earned a nearly perfect score on a high school AP biology exam, it was asked, *What do you say to a father with a sick child?* The chat bot's response astonished Bill Gates, who was involved with ChatGPT and is keen on using artificial intelligence (AI) to address pressing global issues, such as climate change and childhood mortality.

The Microsoft founder-turned-philanthropist wrote in his blog, GatesNotes, "I knew I had just seen the most important advance in technology since the graphical user interface…. It will change the way people work, learn, travel, get health care, and communicate with each other. Entire industries will reorient around it…." He added, "It also raises hard questions about the workforce, the legal system, privacy, bias, and more."

Researchers in the UC Santa Barbara Computer Science Department, and their colleagues both in and beyond the College of Engineering are at the forefront of the AI revolution, pursuing research to develop, refine, examine, and question the rapidly evolving technology while ensuring that it serves all people and harms none. That is why, says computer science professor and AI natural language expert **William Wang,** "We established the Center for Responsible Machine Learning [which has more than sixty faculty affiliates from across campus] to tackle the ethical, legal, and social aspects of AI. It's essential for the AI community to come together and work toward responsible development and deployment."

Cyber defense is another huge area of AI application, and in May (see page 25), the COE received a major National Science Foundation grant to fund a new Institute for Agent-based Cyber Threat Intelligence and OperatioN (ACTION).

In this edition of "FOCUS ON:," we introduce just some of the AI research and researchers at UCSB, acknowledging at the outset that we cannot adequately cover the topic in eight pages and, so, will return regularly in the future to "catch up."



## LANGUAGE, VISION, ACTION

**Yujie Lu,** a second-year PhD student in computer science associate professor **William Wang**'s research group, works on computer vision and natural language processing (NLP), specifically how to implement complex planning problems in autonomous artificial intelligence. Accomplishing that, she says, requires AI models to understand information and signals from three dimensions: language, vision, and action.

"That's how humans do things like making breakfast," says Lu, whose work mostly involves writing code for the application and then running experiments to see if it works. "You need to perceive the environment and have a plan for how to implement the information it gives you, and then you need to be able to take some action."

She has published a first-author paper about language and vision, and another about language and action. A third, expected to be out in late spring or early summer, examines how to use language and vision to initiate action.

The instantly famous large-language model ChatGPT, she says, "is part of the AI that's driving those models, but it doesn't monitor the process it tells you about. It doesn't know what you've done." She hopes to be able to empower ChatGPT to perceive the environment and assess whether a task is being done correctly.

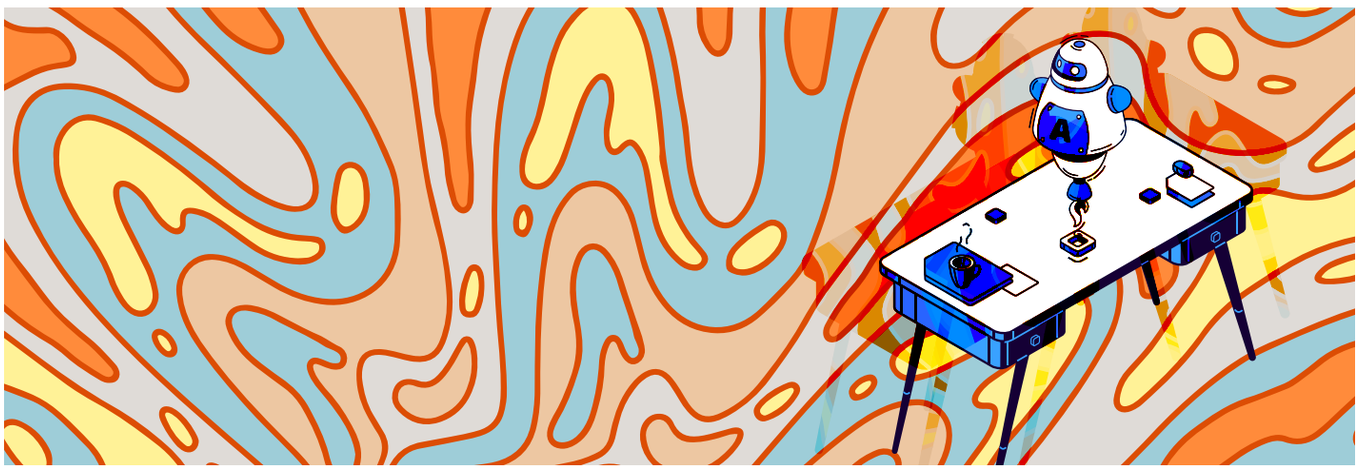
Lu's work and that of others in the field is changing how humans interact with robots. "In the past, when people were trying to train a robot to perform tasks, they had to provide a trajectory, such as a sequence of images reflecting what the robot would see in the environment, so that it could learn to navigate there," she explains. "The robot could then compare the original signals from the environment and the ground truth.

"But if I can use more-natural language, then humans can just give the robot a high-level command like, 'Go to the kitchen.' We can train the robot on those instructions, which are beyond low-level instructions, such as 'Go forward, turn left, turn right.' They have some hierarchy to them. The optimal goal would be for robots to connect a series of high-level natural-language instructions to a series of low-level actions. We want them to operate at a much higher level so that they can learn to reach the goal."

> "THE OPTIMAL GOAL WOULD BE FOR THEM TO CONNECT A SERIES OF HIGH-LEVEL NATURAL-LANGUAGE INSTRUCTIONS TO A SERIES OF LOW-LEVEL ACTIONS... SO THAT THEY CAN LEARN TO REACH THE GOAL."

Progress is occurring. "There is even a robot now that can communicate with people to perform tasks like making coffee or even more complex 'long-horizon' tasks, such as making breakfast," Lu says.

It will be a while, she suggests, perhaps thirty years, before robots exhibit such high-level functioning as being able to follow a simple command, maybe "Go back to the sink and start again," when they get "lost" and can't complete a task. But, she notes, "The technology is evolving very rapidly in a very promising direction."




*Advanced robots and "hallucinating" chatbots add "interest" to the AI landscape.*

## AI AND PLANNING: BUT DO WE WANT *ALL* THAT DETAIL?

**Paul Leonardi,** professor and chair in the UCSB Technology Management Department, essentially studies AI through the lens of "how organizations will operate in the future." In one long-term project started in 2015, he examined how AI-empowered simulations might support decisions in two large metropolitan planning organizations in the U.S. that were trying to develop twenty-year regional plans related to such things as zoning ordinances and where to put transportation corridors.

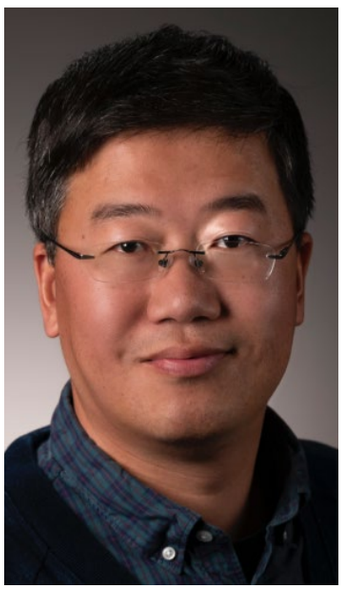
Planners always want more public participation in the process, but Leonardi explains, the models they've used for decades are, essentially, "numbers crunched in an Excel spreadsheet and a couple of bar graphs. They look at traffic congestion and commute time and run a regression analysis." No wonder average citizens at public-comment meetings feel disconnected from the data they don't understand and provide few valuable contributions.

Enter AI, which made it possible to build much more complex models. "We could take all of the data points we have today, run them through these complex algorithms, use AI to stitch the pieces together, make predictions about what's going to happen, render them in three dimensions, and even show a video of what the traffic will look like in an area if certain policies are implemented," Leonardi says. "It's a huge leap forward, but I was curious whether you would get more participation."

The findings were surprising. The two planning organizations made very different choices related to the models. One model was highly detailed; the other was much more sparse. "We thought the more-detailed one would produce much better citizen feedback and input, but it didn't," Leonardi reports. "Instead, when people looked at the model, they said things like, 'Wait, that's my house right there, and I don't want a five-story building around the block from me.' They started complaining about a lot of NIMBY kinds of things, and they didn't talk about the big-picture issues, like *do we want to have a traffic corridor?* Sort of paradoxically, the more detail and precision the model provided, the more the stakeholders focused on the wrong things and not the big picture.

"There's this balance," Leonardi says he learned from the study, "between what the right amount of information is to present and what the right amount of information is to withhold in order to provide the right level of abstraction and effectively stimulate engagement."

Just as we train AI models, we probably will need to train ourselves to work with them.



*Xifeng Yan*

## AUTOMATIC PROGRAMMING: TOWARD DATA DEMOCRATIZATION

"We are on the eve of a big breakthrough in artificial intelligence, something very similar to the evolution of the internet, because it is foundational technology that can support many, many applications," says UCSB computer science professor **Xifeng Yan.**

Despite the limitations of current large language models (LLMs), Yan sees a huge value in them. "Yes, they hallucinate [i.e., embed plausible-sounding random falsehoods into generated content], and they do 'make up' facts, so they need us to verify what they do," he says. "But we are still in charge. AI helps us to finish a task much faster so that we can save our time for other meaningful things. It is much easier to check the bot's work than it is to come up with it yourself from scratch."

*Automatic programming* — a method that can enable an LLM to "auto-write a program so that we can use natural language to query the data" — is another focus in Yan's lab, and something he describes as "a very big development."

"To query and analyze data, you have to be trained to write code, but if you are, maybe, a biologist and don't know how to write code or don't have funding to have someone else do it for you, you can't leverage the huge amount of data you might have," he says. "But what if you could give natural-language commands to an LLM, and it would translate that into programs that would allow you to use your data to get the results you need?

"Sooner or later, you will see this kind of product in the MS Office suite," he continues. "It will democratize data by liberating people from having to learn programming before they can use their data. Everyone will have the freedom to query, manipulate, and analyze data."

Yan also sees tremendous *educational* value in chatbots. "When people think about Chat GPT and education, they might think fabout plagiarism and other negative things, but there is a lot of potential to use it to scale up teaching," he says. "I'm teaching a class with one hundred students. How do I give personalized feedback to every student? Chatbots can help. They can help teachers to plan better lectures and give customized guidance. They can help non-native speakers, such as international students, improve their writing skills. Some students at UCSB learned programming in high school, but others didn't have that chance. ChatGPT and similar models can help those students catch up.

"Imagine that you don't need to learn programming anymore," Yan concludes. "This is the future, and it's happening right now. We are transitioning away from having to write rigid code and to be able to use natural language to get our work done."



*Nelson Phillips*

## MAKING MEANING: AI AND UPDATING THE TURING TEST

The Turing test was created in 1950 to test a machine's ability to exhibit the equivalent of intelligent behavior. While ChatGPT and other large systems have been able to defeat the original game-based test by "passing" as human in a conversation, "What Turing was getting the machine to do wasn't very sophisticated," says **Nelson Phillips,** a professor in the UCSB Technology Management Department. "It's sort of hiding out and managing not to be identified as a computer, but it's not showing human responsiveness."

Phillips and Mark Kennedy, a professor at London's Imperial College, started thinking about an update to the original Turing test, and created their own game-based version that requires players to categorize things and then convince others that their categorizations are reasonable. The point, Phillips says, is to test an AI model's "generative ability to participate in this thing that humans do when we create meaning by coming to a shared agreement about what something means."

They then wrote what Phillips describes as a "highly speculative" paper titled "The Participation Game" which appeared April 25 in *arXiv*, pitching their game as a better test of machine intelligence (than the original Turing test) and offering thoughts about the implications of some of AI's evolving abilities.

For instance, Phillips notes, "One thing humans do that computers can't is to create new symbols and new things, like the concept of a *selfie*," created when someone in Australia sent one to his friend group and first used the word. "They created a new concept," Phillips says, adding, "Humans are fabulous at this. We do it continuously, inventing new concepts like *influencer* or the *minivan* or the *tablet*, which establish a new category in our brain that has a certain set of characteristics."

If AI systems develop that ability, he says, "then they become really powerful but also really troubling. Imagine if you have a bunch of non-human influencers on the internet creating new ideas and populating our culture and our brains with new ways of thinking about the world. You very quickly go to either *Oh, my God, that would be amazing!* or 1984 and Big Brother and *who's controlling these things?*"

Today's chatbots essentially "take human-generated text and other data and mix and match pieces of them, so they are really just greatly amplifying the activity of some human who is actually thinking of the message," Phillips says. "This would just be giving these systems a goal and telling them to come up with and propagate new concepts. They could then convince people to think about the world in a certain way or not, a capacity that is much more powerful than just being able to participate in communication."

> "IMAGINE A BUNCH OF NON-HUMAN INFLUENCERS. YOU VERY QUICKLY GO TO EITHER 'OH, MY GOD, THAT WOULD BE AMAZING!' OR 1984 AND BIG BROTHER AND 'WHO'S CONTROLLING THESE THINGS?'"

## COMBATTING LLM PIRACY

**Yu-Xiang Wang,** a UCSB computer science (CS) assistant professor and co-director of the Center for Responsible Machine Learning, conducts research on the statistical foundation of machine-learning (ML) algorithms while also addressing real-life concerns, such as security, privacy, and copyright protection in the time of large language models (LLMs) like ChatGPt. "We predict that the privacy concerns around AI, natural language processing (NLP), and ML are going to touch every single person on the planet in a few years," he says.

Related to this, people are using LLMs to generate code, which is both promising and troubling. (See "Automatic Programming, on page 19.) "The algorithms have sometimes generated verbatim code from proprietary software from the nineties," Wang says. "We consider such events the growing pains of LLMs, and we need to come up with technical solutions to such problems to make the transition as smooth and as painless as possible."

In addition to privacy and copyright concerns related to LLMs' training data, Wang also worries about the copyright of LLMs themselves. He notes that companies spend billions of dollars to train their models, and then, as OpenAI did, ship them as APIs. In a process referred to as a *model-stealing attack*, a rogue company can then use a well-developed method called *model distillation* to create a copy of the model from the API and then ship it at a fraction of the cost.

In 2022, Wang, his student **Xuandong Zhao,** and fellow CS assistant professor **Lei Li** (see page 22) co-wrote a paper predicting that practice. "Imagine if a rogue company created this model at a fraction of the cost, shipped it as its own, and hijacked all the traffic at a much lower price," he says. "That is not good for innovation."

To combat pirating, Wang's lab came up with "watermarking" tools that can be used to mark any output from an LLM. Then, if someone trains a new model on top of the original, every piece of information it generates will contain a "backdoor" from the original. "It shows up a little bit in the results," Wang says, "and once you get thousands of results, you can collect them and gather the statistical information to prove with high certainty that the model was derived from the original.

"Variants of these watermarking techniques can be used to distinguish AI-generated content, such as college admission essays, homework submissions, music, art, and computer codes from their human-created counterparts." Wang hopes that such research can help to ease the "growing pains" as LLMs change the world.

## TASK-ORIENTED AI: IT TAKES TRUST

Suppose you need physical therapy (PT) but live far from a city and can see a therapist only once every few months. Between visits, how do you make progress? How do you know if you're doing your exercises correctly? How do you get encouragement and guidance?

In her Human-AI Integration Lab, **Misha Sra**, an assistant professor in the UCSB Computer Science Department who recently received a National Science Foundation Early CAREER Award, is developing multimodal AI models to help address such task-oriented needs. "We're looking at building new types of AI-enhanced tools that can augment human physical abilities," she says.

Creating such AI helpers is a huge challenge, and Sra started with a simpler analogous process: making a cup of coffee. "Cooking is a physical task that embodies many of the challenges that we will encounter in building a tool to augment physical abilities," she says. "It has specific procedures, end goals, mechanisms to track progress, variations in how each person follows the procedures, and potential for unknowns, all of which apply to other physical tasks we've thought about, such as physical therapy or fitness training."

In Sra's lab is a setup with everything needed to make coffee. A user puts on an augmented-reality (AR) headset, and the system provides step-by-step instructions, from measuring and grinding the beans to folding the filter, etc. A user who is unfamiliar with a step can choose to receive verbal instructions (if, say, their hands are dirty) or, in a noisy place, watch a first-person perspective video or view a 3D animation shown on top of the relevant object in physical space. A menu attached to the user's hand or one that is gaze-controlled shows them the various help options.

"So, the AI is helping you in real time, instructing you and correcting your errors. Everything is going great, right?" Sra says. "But then the AI makes a mistake, because maybe there's something in the scene that it hasn't seen before. Maybe it tells you that you measured the beans incorrectly, but you didn't. What happens then?

"In such a scenario it is essential for the user to be informed when AI makes an error," she continues. "This information needs to be presented clearly through the user interface in AR. Additionally, the user may wish to understand why the AI made the error, which means that its thought process must be explained to them. Subsequently, the user must decide how to proceed when an error occurs. All of these are design questions that we're currently exploring. The bigger question we want to answer, however, is how to build trust between the human and the AI, despite the AI making mistakes, and how to design a user interface where the human is in control and not the AI."

> "SO, THE AI IS HELPING YOU IN REAL TIME.... EVERYTHING IS GOING GREAT, RIGHT?...BUT THEN THE AI MAKES A MISTAKE.... WHAT HAPPENS THEN?"

The next step, user studies, will help Sra to evaluate the interface and define the path toward the full system design.

*Far left: Second-year PhD student Arthur Caetano uses an augmented-reality headset to follow the AI's instructions and make coffee for Misha Sra:*



## TOWARD NATURAL-LANGUAGE TRANSLATION

One of UCSB Computer Science assistant professor **Lei Li's** main interests is developing advanced AI for language translation and knowledge reasoning. The current crop of commercial translation tools, such as Google Translate, can interpret around one hundred languages, Lei says. "We want to develop translation technology for a thousand languages, to help people better communicate with each other around the world."

That's hard to do, he says, because of a lack of two main elements: web data to train the AI models on lesser-used languages and an algorithm that works well in "low-resource settings."

"When we collect data on the web, we need parallelism, so, if we have a sentence in English, we need the equivalent sentence in, say, Hindi," Li explains. "To teach the machine, we need to pair the languages, not just translate words. We want to develop translation technology that can work well even with limited data to generalize across a very large number of languages."

That process depends on leveraging similarities in linguistic structures, semantics, and even some words that are similar or the same in different languages. Further, Li says, "The same kind of human knowledge is shared in every language. We want to develop AI to automatically learn this universal representation so that it can be generalized across language, and not just for text, but also for speech, so that spoken language can be translated directly to another language."

Over the past two years, researchers in Li's lab have gone from starting with about 36 languages to having 450 in their newest work. During an interview, he pulled out his phone and used his app to translate spoken English into Chinese characters on his screen in real time.

*Lei Li*

He also wants to develop models that can reason in a natural way, in human language, rather than in traditional AI, which takes a more symbolic, formal, mathematical approach to reasoning. His goal is to have AI reason in natural language, the way humans do.

Li gives the example of Los Angeles as the largest city in California, a fact that might lead someone to assert that it is also the state capital. "If we ask the model whether that is correct, we want it to be able to reason with that statement in natural language, to look for evidence, which is also obtained in natural language," he says. "It might find text about California or L.A. on Wikipedia and then piece that together to determine that the statement is incorrect: L.A. is not the capital. That's called *factor verification,* and it's very important right now, including with ChatGPT [with its penchant for hallucinations, i.e., embedding plausible-sounding random falsehoods into generated content], because we need to verify that the generated content is faithful and actually correct."


*An AI-enabled buoy monitors the channel for sounds of endangered whales and sends them to ships in nearly real time so that they can slow down.*

## AI IN THE OCEAN

Collisions with container ships are a top cause of deaths of endangered whales in the Santa Barbara Channel and around the world. UC Santa Barbara researchers have long been involved in seeking ways to eliminate those collisions, and now, **Douglas McCauley,** professor in the Ecology, Evolution, and Marine Biology Department and director of the Benioff Ocean Science Laboratory, is working with colleagues to put artificial intelligence on the side of saving whale lives.

"In the environmental science community, we're looking at existential crises coming down on the planet with climate and biodiversity," McCauley says, "and we're excited to bring the power of AI tools to do good into the domain of environmental problem solving."

Working under the Benioff lab's Whale Safe program, McCauley and colleagues at the Woods Hole Oceanographic Institution developed an AI-powered mapping and analysis tool that collects and displays near-real-time whale data. The technology involves what McCauley describes as a "relatively simple algorithm" that was created for Whale Safe and is integrated with an acoustic detector in the channel, which listens constantly for whales. The underwater microphone turns the acoustic profile of the sounds it hears into an image, like sheet music, and then, McCauley says, "The AI matches those sound images with known images of endangered-whale calls." The classification library used to train the AI was built using years of underwater sound recordings.

"The system notes the whales' presence in the area and identifies them in real time as an endangered blue whale, a humpback whale, or a fin whale," McCauley says. "That information goes to the ships, letting them know that endangered species are in the ocean roadways, so that they can slow down to avoid running over them."

## TASKS, EMPATHY, AND ETHICS IN AI

"I think this is even bigger than the internet," says computer science professor **William Wang** in speaking about ChatGPT and other large language models (LLMs). "It offers a better way to access all the information we have on the internet. It's going to significantly change human society."

One focus of Wang's group is *multimodal AI* — AI that can understand and represent information in non-text-only representations, such as a table, a database with text, image data, and video data. Researchers in his lab are working to improve LLMs to be able to perceive and understand language in its many modes.

As director of the Center for Responsible Machine Learning (CRML) and the Duncan and Suzanne Mellichamp Chair in Artificial Intelligence Design," Wang is especially focused on the ethics of AI, which includes addressing people's fears and concerns about how it might evolve. One big issue in that realm is how to distinguish facts from misinformation.

"LLMs are trained on a gigantic amount of data, so when generating output, they still need humans to verify the truthfulness of the results," Wang explains. "For instance, if you ask an LLM to read a financial statement, it is likely to give you hallucinations [random instances of plausible-sounding, but false, information]. So, for instance, a company's earnings might be $1.2 billion, but the model tells you that rhey are $1.5 billion — a 25 percent error. Those numbers have to be precise and correct, but the models answer questions on numerical reasoning with only about 60 percent accuracy. We still need humans to verify the truthfulness of results from any large generative model." That lack of precision, Wang says, creates a gap between what LLMs can do now and what they might be capable of with significantly increased precision.

Wang is also working to address the fact that AI currently "lacks emotion or empathy." He gives the example of a "failure case" in which Microsoft's search tool Bing asked users to apologize, which, he says, "is really weird. The model doesn't understand pragmatics and social dynamics." For that reason, he says, "We started working on empathetic AI and conversational AI six years ago and wrote a paper about trying to generate an emotional response, a functionality that could be useful in, for example, customer call centers when you want to deploy GPT technologies and ensure that the responses the AI generates are appropriate to the context of a request."

So, while large language models bring many opportunities, Wang notes, "They also present challenges, and we have to figure out how to mitigate them by building more-robust models to reduce this kind of hallucination or misinformation. ChatGPT is a nice demo, but it can't be deployed in an actual product, because those have to be very precise. There is still a lot of fundamental research we have to do to improve these LLMs."

*William Wang*

## SLEUTHING CHATGPT SURPRISES

Ask ChatGPT a question — how to make biscuits or change the oil in your car, or why the sky is blue — and in about twenty seconds a sensible, coherent response will appear on your screen. *Most* of the time, that is, because, as we now know, the model occasionally, infamously *hallucinates* by inserting plausible-sounding random falsehoods into generated content.

While each new version of the model is said to be less biased than its predecessors and also less likely to make up facts, it will still hallucinate, and the company that developed it, OpenAI (now part of Microsoft), continues to provide this warning to users: "Great care should be taken when using language-model outputs, particularly in high-stakes contexts."

"This model is very good at retrieving information and explaining it to people, but it also makes up things and does it in a very convincing way," says second-year PhD student **Xinyi Wang.** "And though it doesn't — and can't — do that intentionally, it can still mislead people, so mitigating that is important to avoid the many potential risks of putting LLMs [large language models] into real-world applications."

In Wang's research, she is "taking a scientific approach to understanding what the model can do and why it can do it." For instance, she notes, "Some of ChatGPT's ability doesn't align with the pre-training objective." She cites an example of something called *in-context learning,* which refers to the fact that models usually learn only during training, but as a model gets bigger, it sometimes seems to learn at what's called *interference time,* and then performs a task that was not part of its training.

*Xinyi Wang*

Wang describes one example in which a model was given four sentences in French and four corresponding English translations. When given a fifth sentence in French *without* the corresponding English translation, the model produced a translation on its own anyway, seeming to have learned from the previous examples.

"That's one weird example, because we did not explicitly train it to generalize to a new instance based on some examples," she says. "So we're studying why that happens."

She says that the most powerful thing about LLMs is this *emergence capability,* when something like an unexpected new skill emerges, which can happen when the model's data size grows. "One thing I want to know is, if we understand how this ability emerges, can we observe other such abilities and make better use of them?" she says. "Also, as we better understand the model, we hope to be able to mitigate any negative effects of such events."



## BUT DOES IT KNOW ME?

Doctors sometimes don't know, or remember, much about us. While changing his primary-care physician, UC Santa Barbara's director of the Center for Information Technology and Society, social scientist **Joe Walther,** started thinking about whether he would be better "known" by a doctor who might have only his latest lab test results, or an AI system that "can memorize my medical history in no time and conjure up whatever information is relevant."

Out of that arose the question: what does it take to feel known — by another human and by an AI system? To find out, Walther collaborated with Penn State University professor S. Shyam Sundar and his students, who ran experiments in which human subjects did an online intake interview after being told they were interacting with a human doctor or an AI system. In the scripted dialog, the doctor entity would ask questions about exercise, diet, etc. In a follow-up a week later, two conditions were added: the doctor either remembered information from the previous interaction or didn't appear to remember it at all. When the human or AI doctor didn't recall data from the prior encounter, it would repeat previously asked questions, such as, "Are you getting any exercise?"

People in the study liked it when the human doctor remembered their history, but found it "creepy" when the AI did; they were worried about where the remembered information was going.

"So, then we asked ourselves what it means to know someone, beyond their medical information," Walther says. "We did the experiment again, but this time added that the doctor either remembers social information about the patient or doesn't. It might ask during the first visit, "How do you like to be addressed?" and use that form of address the next time, and say such things as "Last time, you mentioned you have a good relationship with your family. Is that still the case?"

In this instance, people liked both the AI and the human doctor better when they remembered and brought such information into the conversation. Under that condition, they also liked the AI's recall of their medical information. "So, if the AI remembered social stuff, they also liked that it remembered medical stuff," Walther says.

"We used to think that to be known by somebody was to be remembered by them in a distinctive way," Walther notes. "Now, we have to refine that. For AI to make you feel that it knows you, it has to remember personal and social aspects, not just the task-oriented data. It's still kind of surprising to me that, even though it's a machine that we recognize as being a machine, we like it better if it's 'personable' with us."

*Joe Walther*

## USING AI TO MAP METHANE EMISSIONS

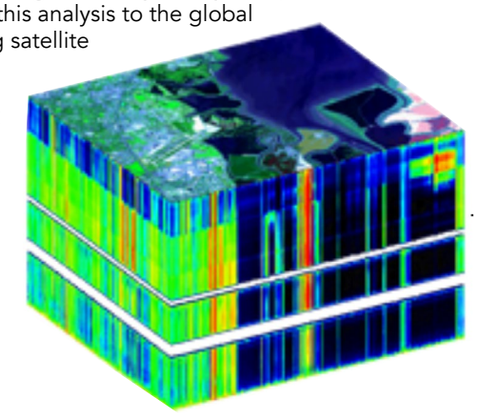
At the 2022 World Climate Summit, methane emissions were identified as being responsible for 30 percent of Earth's warming. "To put that into perspective," says **Satish Kumar,** a fifth-year PhD student in the Vision Research Lab (VRL), led by **B. S. Manjunath,** distinguished professor and chair of the Electrical and Computer Engineering Department at UC Santa Barbara, "The amount of damage to the environment that $CO_2$ will do in 100 years, methane can do in just 1.2 years." The problem has worsened in the U.S. since about 2010, when domestic oil and natural-gas production exploded.

Manjunath, a pioneer in the field of big-image-data management and director of the Center for Multimodal Big Data Science and Healthcare, was recently elected to the National Academy of Inventors (NAI). In the realm of multimodal big data, he developed the open-source BisQue (Bio-Image Semantic Inquiry User Environment) image informatics platform, the intent of which was, he says, "to enable reproducible computer vision [a main topic of AI research], bringing together data, annotations, and methods so that researchers could reproduce their results at anytime.

To help address the methane problem, Kumar developed a computer-vision/AI model, called the MethaneMapper (the data and AI models are distributed through the BisQue software platform), to detect emissions from two key sources: the gas and oil sector, which make up 36 percent of total methane emissions in the U.S., and agriculture and dairy farms, which account for 26 percent. He is the lead author on a scientific paper about the MethaneMapper that will be featured as a "Highlight Paper" at the 2023 Computer Vision and Pattern Recognition (CVPR) Conference, the premiere event in the computer-vision field, to be held this June in Vancouver, British Columbia.

Several methane-monitoring systems are currently in place, but existing methods for analyzing those images are not scalable, because they are prone to significant error and must be inspected by domain experts. MethaneMapper does everything automatically.

Kumar proposes what the paper describes as "a novel end-to-end spectral absorption wavelength aware transformer network…that introduces novel modules to help locate the most relevant methane plume regions." One module transforms what is called a *HyperSpectral data cube* into an aerial map displaying the location of a methane plume on the ground. Says Manjunath, "Our vision is to scale this analysis to the global level, using satellite images."



*The stacked images of a HyperSpectral data cube reveal specific elements on Earth, such as methane plumes.*

## FASTER, MORE-EFFICIENT, MORE-AFFORDABLE AI

Given their remarkable abilities, perhaps its not surprising that the new AI-driven large language models (LLMs), like ChatGPT, are, well, power hungry. Training and running them is energy-intensive and expensive — the initial release of ChatGPT was trained on ten thousand Nvidia GPUs, at $1,500-$2,000 per unit, or $20 million — which is one reason why, for now, LLMs are the exclusive realm of large companies. UCSB Computer Science assistant professor **Yufei Ding** is working to make LLMs faster, more customizable by individual users, less expensive, and more energy-efficient as a way of reducing their carbon footprint.

Ding conducts research on three main fronts. The first is designing a hardware accelerator tailored for LLM computing, unlike the CPU in a laptop or desktop computer, which, she says, "is tailored for more 'general' computing. "It's a change in the architecture of the hardware itself."

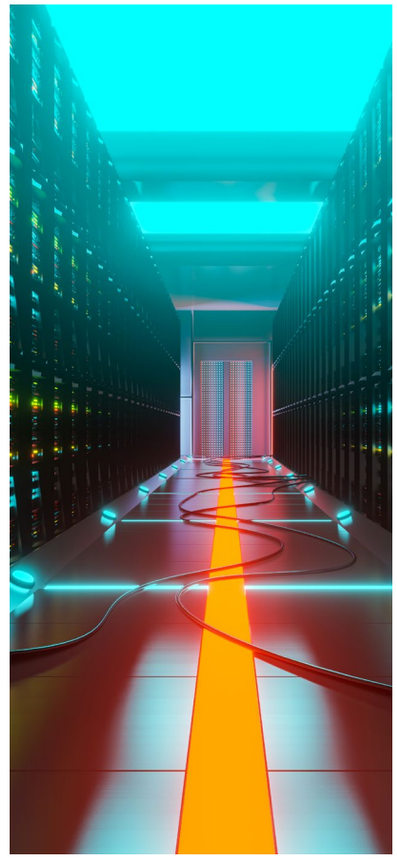
The second area is software optimization, where she seeks to ensure that, "As hardware gets more and more complicated, an application can utilize it optimally, automatically," Ding says.

The third area is fine-tuning of the models. "At the algorithm level, instead of doing end-to-end training [complete training of the model] for everything, maybe we can have a general, powerful foundation model that we need to train once and that will just need some lightweight fine-tuning, such as tuning ChatGPT for medical care — to use it for other applications later," Ding explains.

"Or maybe I want to give personal information to ChatGPT so that it can help revise my paper, but I want to keep it private; I don't want the model to be trained on my data," she explains. "That's a fine-tuning process that could be done only on my own computer. Big companies have many thousands of GPUs running together, but I might have only one single laptop. How can I do that fine-tuning? It puts new challenges on the hardware and software designs."

The various areas of Ding's work address different scales of optimization that grow in scope and layer upon each other, from the smallest, a single device, to multiple devices within a node, up to inter-node coherence and communication. For end-to-end training, big companies are most concerned with parallelizing their thousands of servers to optimize efficiency and service. For a small company or an individual trying to fine-tune an LLM, privacy might be the main concern.

"Things like what kind of hardware you have, what you can afford, and what kind of task you want to do determine the optimization you need to have," she says. "We want to work across scales to ensure good performance in all kinds of scenarios."



*Yufeig Ding wants to make everything about AI — including server centers like this one (right) — more energy efficient.*

## UCSB WILL LEAD A NSF COLLABORATION TO DEVELOP AI FOR NATIONAL INFRASTRUCTURE DEFENSE

"Computer systems are increasingly central to national infrastructure in the financial, medical, manufacturing, defense, and other domains. That infrastructure is at risk from sophisticated cyber-adversaries backed by powerful nation-states whose capabilities rapidly evolve, demanding equally rapid responses."

That passage is taken from the abstract for the ACTION Institute, which the National Science Foundation has just established to develop advances in "artificial intelligence and autonomous reasoning that will be tightly integrated with advanced security techniques to identify and correct vulnerabilities, detect threats and attribute them to adversaries, and mitigate and recover from attacks."

*ACTION Institute Logo*

UC Santa Barbara is the lead institution on the $20 million, five-year NSF grant, and UCSB computer science professor and cyber-security expert **Giovanni Vigna** is the director. The project links UCSB and ten collaborating universities — UC Berkeley, the University of Washington, the University of Illinois Chicago, the University of Illinois Urbana-Champaign, the University of Chicago, Purdue University, Rutgers University, the Georgia Institute of Technology, Norfolk University, and the University of Virginia.

Together, researchers at those institutions "will develop novel approaches that leverage artificial intelligence — informed by and working with experts in security operations — to perform security tasks rapidly and at scale, anticipating the moves of an adversary and taking corrective actions to protect the security of computer networks as well as people's safety. The Institute will function as a nexus for the AI and cybersecurity communities, and its research efforts will be complemented by innovation in education from K-12 to post-doctoral students, the development of new tools for workforce development, and the creation of new opportunities for collaboration among the Institute's organizations and with external industry partners."

The institute will initiate a revolutionary approach to cybersecurity, in which AI-enabled intelligent security agents cooperate with humans across the cyber-defense life cycle to jointly improve the security posture of complex computer systems over time.

*This was late-breaking news. Be sure to watch for more in-depth coverage of the institute in a future issue of* Convergence.