Zuckerberg says phones are nearly obsolete

In a bold proclamation, Mark Zuckerberg has declared that the era of smartphones is nearing its end, with smart glasses and augmented reality poised to take their place. This revelation marks a significant shift in how we perceive and interact with technology, hinting at a future where mobile phones may become relics of the past.

The Vision of a Post-Phone World

Mark Zuckerberg's announcement about the impending obsolescence of smartphones has sparked considerable interest and debate. His vision centers on the rise of augmented reality (AR) and smart glasses, which he believes will take over as the principal means of communication and interaction with digital content. This transition aligns with Meta's broader goals of creating a more immersive and interconnected digital ecosystem, often referred to as the metaverse.

Augmented reality and smart glasses are set to become the successors to smartphones, offering a seamless blend of digital and physical worlds. By

overlaying digital information onto the real world, these technologies promise to enhance our daily interactions, making information more accessible and experiences more engaging. Zuckerberg's vision is not just about technology for technology's sake; it's about creating a more fluid and natural interaction with our digital lives.

The Rise of Smart Glasses

Technological advancements have made the concept of smart glasses not only feasible but also intriguing. Improvements in AR technology, battery life, and miniaturization of components are key factors driving this progress. Smart glasses are being designed to offer a hands-free experience, with features such as voice control, gesture recognition, and integrated AI, setting them apart from current mobile devices.

Despite the promising features, the adoption of smart glasses as mainstream technology faces several challenges. These include the need for significant improvements in display technology, battery efficiency, and user interface design. Moreover, ensuring a comfortable and stylish design that users are willing to wear daily is crucial for widespread adoption. However, the opportunities are immense, as smart glasses have the potential to revolutionize various sectors, from healthcare to retail.

Implications for Consumers

The transition from smartphones to smart glasses could have profound effects on daily life and social interactions. With information readily available in one's field of view, tasks could become more efficient, and communication more immediate. This shift could also redefine social norms and etiquette, as the presence of wearable technology might change perceptions of privacy and attention in social settings.

For consumers, the benefits of embracing smart glasses include enhanced convenience and new forms of interaction with the digital world. However, there are potential drawbacks, such as concerns over privacy and security. The ability of smart glasses to record and analyze one's surroundings raises

significant privacy issues, necessitating robust security measures and clear guidelines to protect user data.

Impact on the Tech Industry

As the tech industry prepares for this paradigm shift, major companies, including Meta, are investing heavily in AR and smart glasses technology. This is not only a move to stay competitive but also to lead the charge in defining the next generation of personal technology. The competitive landscape is heating up, with other tech giants developing their own AR solutions, aiming to capture a share of this emerging market.

The potential economic implications for the smartphone manufacturing sector are significant. As consumer preferences shift, companies may need to pivot their strategies, invest in new technologies, and retool manufacturing processes. This evolution presents both risks and opportunities, as firms that adapt swiftly could gain a competitive advantage in the burgeoning AR market.

Sociocultural Shifts

The advent of smart glasses is likely to impact communication habits and digital etiquette. As AR becomes more integrated into daily life, the way individuals interact with each other and their environment could change dramatically. This technology has the potential to reshape sectors such as entertainment, education, and work, offering new possibilities for immersive learning and collaboration.

However, sociocultural resistance and acceptance will play a crucial role in the adoption of smart glasses. As with any transformative technology, there will be those who embrace it enthusiastically and those who are more cautious, concerned about issues such as privacy, security, and the impact on human relationships. Understanding and addressing these concerns will be key to facilitating a smooth transition to a post-phone era.

Future Prospects and Challenges

While predicting the exact timeline for the widespread adoption of smart glasses is challenging, many experts believe it could happen within the next decade. For this technology to succeed, several key challenges must be addressed, including improving AR displays, ensuring long battery life, and creating intuitive user interfaces.

The role of developers, designers, and policymakers will be critical in shaping the future of wearable technology. Developers and designers must focus on creating compelling applications and experiences that leverage the unique capabilities of smart glasses. At the same time, policymakers need to establish regulations that safeguard user privacy and security while fostering innovation.

Mark Zuckerberg, July 30, 2025

Personal Superintelligence

Over the last few months, we have begun to see glimpses of our AI systems improving themselves. The improvement is slow for now, but undeniable. Developing superintelligence is now in sight.

It seems clear that in the coming years, AI will improve all our existing systems and enable the creation and discovery of new things that aren't imaginable today. But it is an open question what we will direct superintelligence towards.

In some ways, this will be a new era for humanity, but in others it's just a continuation of historical trends. As recently as 200 years ago, 90% of people were farmers growing food to survive. Advances in technology have steadily freed much of humanity to focus less on subsistence and more on the pursuits we choose. At each step, people have used our newfound productivity to achieve more than was previously possible, pushing the frontiers of science and health, as well as spending more time on creativity, culture, relationships, and enjoying life.

I am extremely optimistic that superintelligence will help humanity accelerate our pace of progress. But perhaps even more important is that superintelligence has the potential to begin a new era of personal empowerment where people will have greater agency to improve the world in the directions they choose.

As profound as the abundance produced by AI may one day be, an even more meaningful impact on our lives will likely come from everyone having a personal superintelligence that helps you

achieve your goals, create what you want to see in the world, experience any adventure, be a better friend to those you care about, and grow to become the person you aspire to be.

Meta's vision is to bring personal superintelligence to everyone. We believe in putting this power in people's hands to direct it towards what they value in their own lives.

This is distinct from others in the industry who believe superintelligence should be directed centrally towards automating all valuable work, and then humanity will live on a dole of its output. At Meta, we believe that people pursuing their individual aspirations is how we have always made progress expanding prosperity, science, health, and culture. This will be increasingly important in the future as well.

The intersection of technology and how people live is Meta's focus, and this will only become more important in the future.

If trends continue, then you'd expect people to spend less time in productivity software, and more time creating and connecting. Personal superintelligence that knows us deeply, understands our goals, and can help us achieve them will be by far the most useful. Personal devices like glasses that understand our context because they can see what we see, hear what we hear, and interact with us throughout the day will become our primary computing devices.

We believe the benefits of superintelligence should be shared with the world as broadly as possible. That said, superintelligence will raise novel safety concerns. We'll need to be rigorous about mitigating these risks and careful about what we choose to open source. Still, we believe that building a free society requires that we aim to empower people as much as possible.

The rest of this decade seems likely to be the decisive period for determining the path this technology will take, and whether superintelligence will be a tool for personal empowerment or a force focused on replacing large swaths of society.

Meta believes strongly in building personal superintelligence that empowers everyone. We have the resources and the expertise to build the massive infrastructure required, and the capability and will to deliver new technology to billions of people across our products. I'm excited to focus Meta's efforts towards building this future.

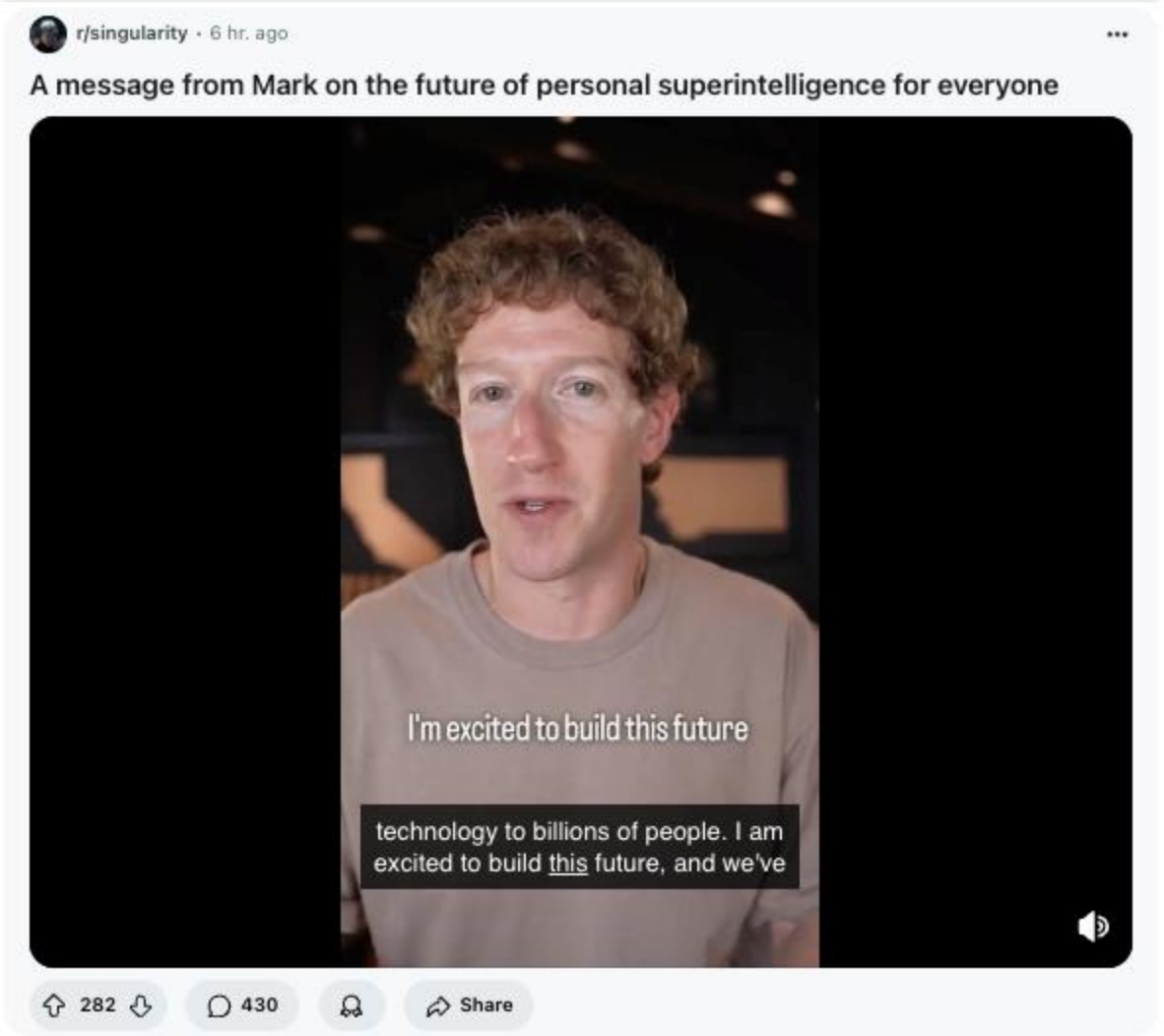
Meta's $72B Superintelligence Bet is Building a Future That Could Kill Its Own Apps

Meta's spending $72 billion to build "personal superintelligence" that'll free us from our screens—while their current AI just made us 20% more addicted to Instagram—but we think true "personal" superintelligence will kill Facebook

and Instagram, and what they're actually doing is racing to build the replacement before someone else does.

Grant Harvey, July 30, 2025

Good ole Marky Mark over at Meta just pitched his vision for "personal superintelligence": AI that'll be built specifically for people's personal lives, not their productivity software.



Mark's basically betting on a world where we spend less time working and more time creating or chatting with our AI powered glasses.

TBH, he's probably not wrong about that…

We here at The Neuron think the #1 most positive thing AI will be able to do for us in the future, without question, is get us away from sitting at desks all day staring at screens.

But what does that mean for Facebook, and Instagram?

After all, if we're not glued to our phones scrolling Instagram, how does Meta make money?

The $47.5 Billion Irony

Here's the wild part: While Zuck's dreaming about getting us off our screens, Meta's current AI is doing the exact opposite—and printing money in the process.

The Q2 earnings call revealed some eye-popping numbers:

- Instagram video time shot up 20%+ year-over-year.
- Better AI content matching saw time spent on Facebook increase by 5%, and Instagram increase 6%—just this quarter.
- AI-powered ad improvements drove 5% more conversions on Instagram, 3% on Facebook.

- Revenue hit **$47.5 billion** (up 22% YoY), with ads doing most of the heavy lifting.

Second Quarter 2025 Financial Highlights

In millions, except percentages and per share amounts	Three Months Ended June 30, 2025	2024	% Change
Revenue	$ 47,516	$ 39,071	22 %
Costs and expenses	27,075	24,224	12 %
Income from operations	$ 20,441	$ 14,847	38 %
Operating margin	*43 %*	*38 %*	
Provision for income taxes	$ 2,197	$ 1,641	34 %
Effective tax rate	*11 %*	*11 %*	
Net income	$ 18,337	$ 13,465	36 %
Diluted earnings per share (EPS)	$ 7.14	$ 5.16	38 %

Apparently, Meta's AI ranking systems are so good at showing us content we can't resist that "more than two-thirds of Instagram's recommended content in the U.S. is now original posts"—meaning Reels isn't just recycled TikToks anymore. It's becoming a legitimate TikTok competitor.

How do they do this? In part, by optimizing for our interest *in the exact moment we're using the app.* This is powerful stuff; the better AI gets at giving us what we want, when we want it, without giving us things we DON'T want when we're focused on something else, the longer we'll engage.

The company's even rolling out AI tools that let advertisers generate videos, translate captions into 10 languages, and create campaigns that basically optimize themselves. Nearly 2 million advertisers are already using these features, and seemingly they're working, because said advertisers keep buying ads.

So right now, "regular" AI is Meta's cash cow... and its keeping us glued to our feeds longer than ever.

The $72 Billion Plot Twist

Now, here's where it gets spicy: Meta's planning to spend up to $72 billion on "superintelligence" infrastructure in 2025—that's a $30 billion increase from this year. And they're not stopping there. CFO Susan Li basically said "buckle up, we're doing this again in 2026."

As we've shared before, they're building massive AI clusters with names straight out of Greek mythology:

- **Prometheus** in Ohio: Coming online in 2026 as the world's first gigawatt+ cluster.
- **Hyperion** in Louisiana: Could scale to 5 gigawatts (Manhattan-sized footprint, btw).
- Plus "multiple more Titan clusters" they haven't even named yet.

All those data centers aren't just to make Facebook's algorithms better. Why do you think the Zuck is laying down Scrooge McDuck-sized compensation packages to poach AI researchers from all the big AI companies? He's trying to YOLO his way to Superintelligence. Which brings us back to his original pitch.

WTF is "Personal Superintelligence" Anyway?

Breaking down Zuckerberg's vision for "personal superintelligence," here's what he's actually describing:

What it does:

- Helps you achieve your personal goals.
- Lets you "create what you want to see in the world."
- Enables you to "experience any adventure."
- Makes you "a better friend to those you care about."
- Helps you "grow to become the person you aspire to be."

How it works:

- AI glasses that see what you see and hear what you hear.
- Interacts with you throughout the day as your "primary computing device."
- "Knows us deeply" and "understands our goals."
- AI systems that can improve themselves (he says they're already seeing "glimpses" of this).

In practicality, this would probably work like a local router that interfaces with whatever cloud models, services, and agents you run on a day to day basis to help live your life; like an always on secretary, coordinating between all these various feeds and inputs. *That's us drawing some of our own conclusions, though.*

The bigger philosophy:

- *Individual empowerment* vs. centralized automation.
- Against the model where AI does all the work and "everyone lives on a dole of its output" (*this is a direct shot at Sam and Dario*).
- Less time in productivity software, more time "creating and connecting."
- Continuation of the historical trend where technology frees people from survival tasks so they can pursue what they actually want.

That last point sounds pretty good to us, NGL...

What makes it different from competitors:

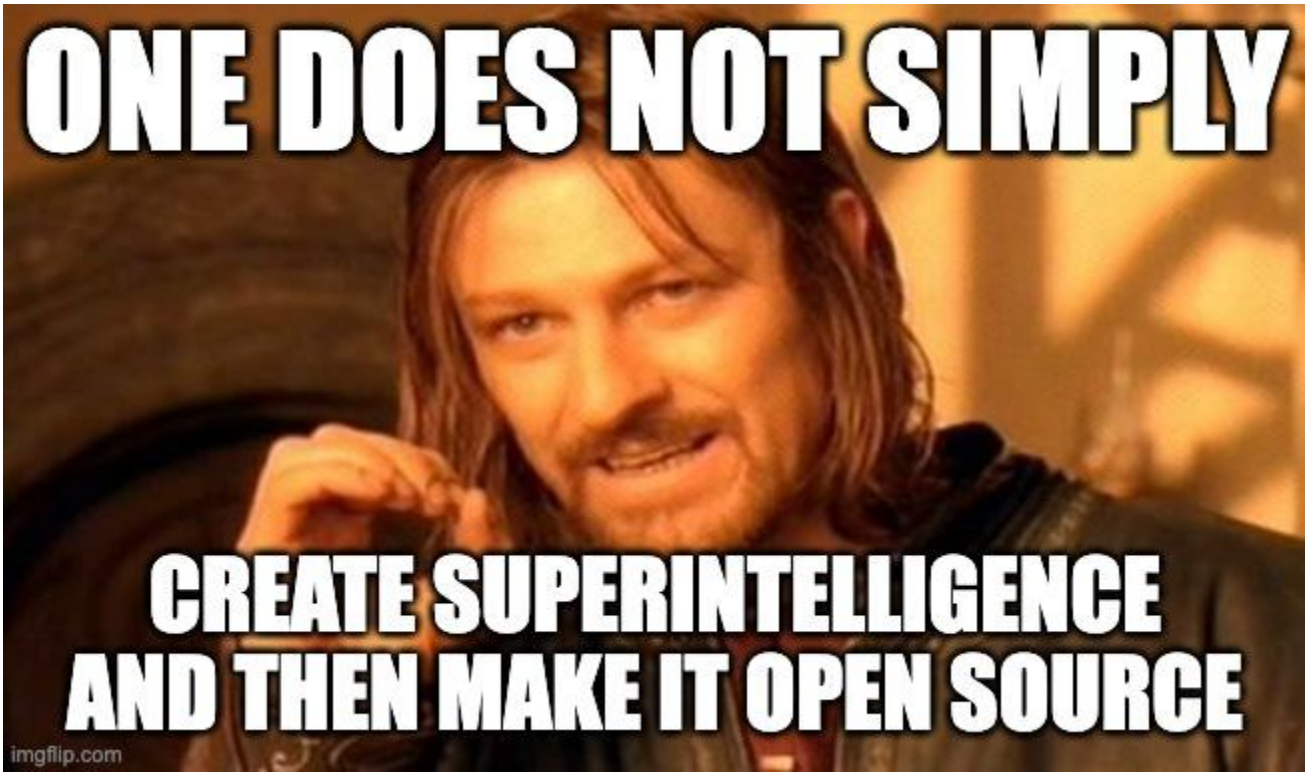
- Others are building AI to "automate all valuable work" centrally.
- Meta wants to put superintelligence power directly in individual hands.
- Focus on amplifying human aspirations rather than replacing humans.

The key insight: He's essentially describing a hyper-personalized AI assistant that lives in your glasses, knows everything about you, and helps you become the best version of yourself—rather than an AI that just does your job for you.

It's ambitious as all get out, but also pretty vague on the actual mechanics. The closest real-world example might be if Siri could see through your eyes, had perfect memory of your life, and was 100x smarter at helping you achieve whatever you wanted.

Now, Mark's "power to the people" messaging would all check out fine and good with how Meta used to do AI development (a.k.a releasing all of its models open source; arguably, they set the standard for open source releases, and the industry owes them a HUGE debt of gratitude for that).

But one does not simply create superintelligence and make it open source



Since there's a chance not everything will be open and local (because how could it be??), there's still going to be some element of centralized AI at play here. And if you check the Reddit comments on Mark's pitch, a lot of folks don't trust Meta with their data now, so it's hard to imagine they'll trust Meta with their data when they're literally streaming their entire life all the time via smart glasses. *Then again, folks on Reddit are likely not FB and IG stans, hence them being on Reddit. i*

Speaking of glasses...

The Glasses-Shaped Elephant in the Room

As reported by TechCrunch's Sarah Perez, during the earnings call, Zuck dropped this beauty: "I think in the future, if you don't have glasses that

have AI — or some way to interact with AI — I think you're … probably [going to] be at a pretty significant cognitive disadvantage compared to other people."

He's basically saying not having AI glasses will be like showing up to a gun fight with a butter knife. Except instead of a duel, it's an all out, knock down, drag out 24/7 brawl vis-a-vis the Running Man because it's your life all the time wherever you go.

As reported, Ray-Ban Meta glasses are already seeing sales 3x year-over-year (proof that a significant cohort doesn't care about Mark's stewardship over their data). Also, Meta AI previousl reported having a billion monthly users, and estimated 3.4B people across FB, Instagram, Messenger, and WhatsApp used it daily.

So yeah, Zuck's convinced that "glasses are basically going to be the ideal form factor for AI because you can let an AI see what you see throughout the day, hear what you hear, talk to you." And having that advantage will dramatically speed up your day and improve your life...so much so that it could be seen as an advantage against those who hold out or can't access it.

What Social Media Looks Like Through Glasses

But what does this actually mean for Facebook and Instagram? We see a few possibilities:

Option 1: The Ambient Feed - Imagine viral shorts playing softly in your peripheral vision, like a heads-up display you can ignore or engage with via

eye gestures. Swipe right with your eyes to save for later, look away to dismiss.

Option 2: The AI Curator - Your personal AI agent becomes your feed, occasionally whispering "Hey, your friend just posted something hilarious" or showing you a perfectly-timed meme when you're waiting for coffee.

Option 3: The Living Room - Social becomes spatial. Instead of scrolling through videos, you pop into live, always-on video hangouts with friends. Think Discord voice channels but for your entire visual field—hop in and out of your friends' "rooms" whenever you're bored.

Whatever form it takes, Meta will need to build an entirely new platform from the ground up. You can't just port Instagram to AR—that's like trying to put a website on a smartwatch. The whole interaction model needs reinventing.

The Beautiful Contradiction

This creates what we're calling a "beautiful contradiction." To us, it's obvious, but nobody seems to be talking about it:

Meta is simultaneously building two futures that can't coexist.

- **Future #1:** Today's Meta, where AI makes our feeds so addictive we spend 20% more time watching Instagram videos.
- **Future #2**: Tomorrow's Meta, where we're living our best lives through AR glasses, spending LESS time in "productivity software" (which, let's be honest, includes scrolling Instagram).

So as we "cut the cord" of screen-based work and switch to this ambient, augmented way of living, what happens to the attention economy that funds Meta's entire empire?

The Real Play

We think Meta knows exactly what they're doing. They're not building superintelligence to enhance Facebook—***they're building it to replace Facebook.***

When AI handles more of our work, we'll naturally want to spend more time doing real-world stuff. That means the shrinking amount of time we *do* spend on digital interactions needs to be incredibly valuable and efficient.

Think about it: If we're all wearing AR glasses and living more "IRL" as Zuck predicts, the entire interaction paradigm changes. Instead of infinite scrolling, we'll need what we're calling "*impactful interactions*"—shorter, more precise digital moments that don't pull us out of real life.

Meta won't monetize through banner ads in your peripheral vision (nobody's keeping those glasses on). Instead, they'll likely create something like "sponsored transactions"—turning advertising into gas fees for the metaverse rather than interruptions in your feed.

It's like Netflix building streaming while still mailing DVDs. They know which way the wind is blowing, and they're spending $72 billion to make sure they're the ones blowing it.

Sam Altman already has the 1.0 answer for this in construction for OpenAI—he's planning to charge a 2% commission when people buy stuff through ChatGPT. Zuck could easily do the same thing.

Imagine it: Your AI glasses help you buy concert tickets, and Meta quietly takes a cut. No need for banner ads cluttering your view (*because let's be real, people would rip those things off immediately*), just tiny transaction fees on everything you do.

Now that stablecoins and crypto have begun to get legitimized through US legislation, Meta could even bring back its long dead Diem currency (which has quite the story, #GoneTooSoon) in order to cut Visa and Mastercard out of the equation (*not everyone can take a cut in this world, or it'll be death by a thousand cuts*).

Here's the problem though—as established, people *really* don't trust Zuckerberg with this kind of power.

It's a smart business model. Whether people will actually trust Meta with their personal AI assistant? *That's the $1.76 trillion question.*

The one thing everyone who's paying attention does agree on is that this battle for superintelligence is existential for Meta. Like Zuck said on the call, the combination of AR glasses and AI could even redeem the Metaverse idea, which again, was probably a few years too early, but in a few years from now, doesn't seem that ridiculous (*in a post superintelligence world for example, should one materialize*).

The point is, more free time does NOT by default equal more digital time. To us, it should mean LESS digital time. If you're working less, you won't want to hang out with your friends more on Facebook. You'll want to hang out with them more in France, or Bali.

So no, personal superintelligence will not mean more digital time, but it could mean more *augmented* time. More *hybrid* time, time spent having hybrid experiences where AI and augmented reality work together to bring the convenience of the online world into the physical world to bring little joys or insights to us wherever we might go. To *accentuate reality*, not keep us from it. Therefore, Mark's best bet is to free us from the stationary screen and welcome us back into the real world.

The Bottom Line

Meta's betting that superintelligence won't just change *how* we use technology—it'll change *why* we use it. Today's AI keeps us scrolling. Tomorrow's AI will (allegedly) set us free.

The wild part? They might actually pull it off. With billions of people already using Meta AI (in some form or another), glasses that are actually selling, and enough compute power to run a small country, Meta's positioned itself to own the transition from screens to... whatever comes next.

Just don't expect them to tell Wall Street they're planning to nuke their own business model. Some revolutions are better kept quiet—at least until you've built the gigawatt clusters to power them.

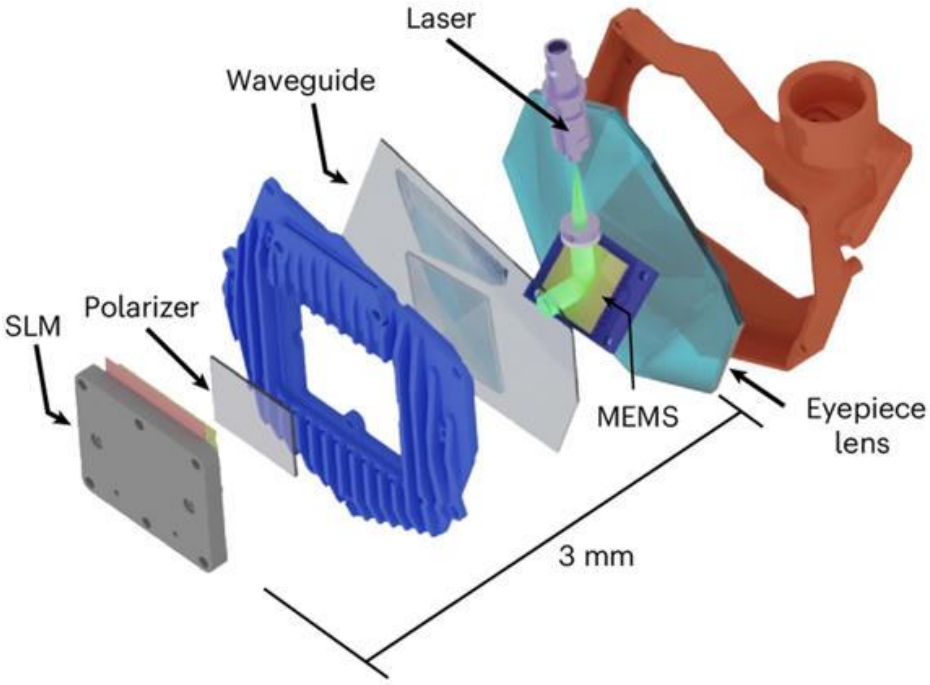
Meta just revealed a new XR display prototype — and it may be the biggest leap in smart glasses since Google Glass

Smart glasses are evolving at a rapid pace — incredible display tech coming to AR specs and AI breakthroughs making the likes of the Oakley Meta HSTN's incredibly intelligent. But we all know the end goal is a pair of glasses that brings these two worlds together.

And researchers at Meta Reality Labs and Stanford University may have just given us the clearest glimpse yet of this future. Yes, I know the researcher pictured above kind of looks like a Cyberpunk pirate, but that eyepatch is a prototype holographic XR display — creating full 3D holograms on a screen thin enough to be used in a standard-sized pair of glasses.

This is the holy grail that companies have been chasing down, and Meta just took one giant step closer to it.

How does it work?



So let's get into the details of this mixed reality tech. Basically, it's a combination of custom glass and silicon along with AI-driven algorithms to render "perceptually realistic 3D images," as the researchers say in their paper.

To make this happen, there's a custom ultra-thin waveguide display driving the visuals — using a laser to project onto a uniquely-textured part of the lens glass for picture clarity.

After this, it goes through a polarizer so we can see it, and a custom-designed Spatial Light Modulator that will (you guessed it) modulate light — the special thing being that it will do so on an individual pixel-by-pixel basis to render a "full-resolution holographic" image we can see.

That's right. *Holograms*. Unlike your standard VR headsets and AR glasses that use eye-confusion techniques to simulate depth, this system can produce true holograms by reconstructing them entirely through light.

There's both a wide field of view and a large eyebox to accommodate all possible pupil positions for accessibility.

“The world has never seen a display like this with a large field of view, a large eyebox, and such image quality in a holographic display,” Gordon Wetzstein, Stanford electrical engineering professor, told the Stanford Report.

“It’s the best 3D display created so far and a great step forward – but there are lots of open challenges yet to solve.”

And even better? All of this is crammed into a panel just 3mm thin! That is **significant** for the future progress of stuffing displays into glasses without needing bird bath optics.

The future is in our sights

One challenge to note, though, is that this is mixed reality and not augmented. The wording is critical here, as the mixed reality reference means the optics are **not** transparent.

But that being said, with this being the second of three projects to bring this to life in a commercial project, there’s no doubt that what we’re looking at here is a breakthrough. Don’t expect this tech to come to glasses you can buy for another few years, but if this mixed reality display can become augmented, that’s the dream I’ve been having over the past four years of testing and writing about smart glasses!

Synthetic aperture waveguide holography for compact mixed-reality displays with large etendu, Published: 28 July 2025

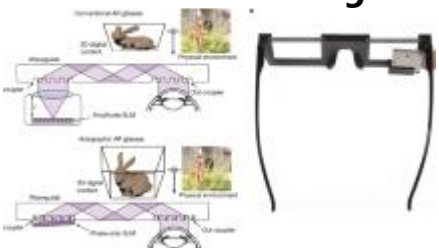
- Suyeon Choi,
- Changwon Jang,
- Douglas Lanman &
- Gordon Wetzstein

Nature Photonics (2025)Cite this article

Abstract

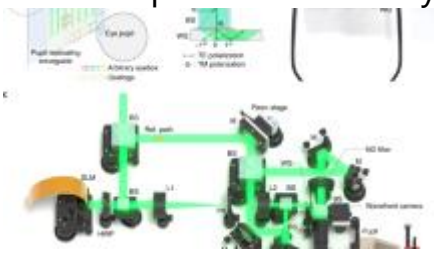
Mixed-reality (MR) display systems enable transformative user experiences across various domains, including communication, education, training and entertainment. To create an immersive and accessible experience, the display engine of the MR display must project perceptually realistic 3D images over a wide field of view observable from a large range of possible pupil positions, that is, it must support a large étendue. Current MR displays, however, fall short in delivering these capabilities in a compact device form factor. Here we present an ultra-thin MR display design that overcomes these challenges using a unique combination of waveguide holography and artificial intelligence (AI)-driven holography algorithms. One of the key innovations of our display system is a compact, custom-designed waveguide for holographic near-eye displays that supports a large effective étendue. This is co-designed with an AI-based algorithmic framework combining an implicit large-étendue waveguide model, an efficient wave propagation model for partially coherent mutual intensity and a computer-generated holography framework. Together, our unique co-design of a waveguide holography system and AI-driven holographic algorithms represents an important advancement in creating visually comfortable and perceptually realistic 3D MR experiences in a compact wearable device.

Similar content being viewed by others



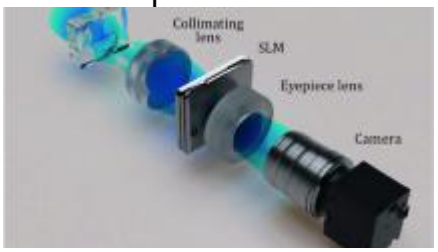
Full-colour 3D holographic augmented-reality displays with metasurface waveguides

Article Open access08 May 2024



Waveguide holography for 3D augmented reality glasses

Article Open access02 January 2024



Expanding energy envelope in holographic display via mutually coherent multi-directional illumination

Article Open access22 April 2022

Main

Mixed reality (MR) aims to seamlessly connect people in hybrid physical–digital spaces, offering experiences beyond the limits of our physical world. These immersive platforms provide transformative capabilities to applications including training, communication, entertainment and education[1,2], among others. To achieve a seamless and comfortable interface between a user and a virtual environment, the near-eye display must fit into a wearable form factor that ensures style and all-day usage while delivering a perceptually realistic and accessible experience comparable to the real world. Current near-eye displays, however, fail to meet these requirements[3]. To project the image

produced by a microdisplay onto a user's retina, existing designs require optical bulk that is noticeably heavier and larger than conventional eyeglasses. Moreover, existing displays support only two-dimensional images, with limited capability to accurately reproducing the full light field of the real world, resulting in visual discomfort caused by the vergence–accommodation conflict[4,5].

Emerging waveguide-based holographic displays are among the most promising technologies to address the challenge of designing compact near-eye displays that produce perceptually realistic imagery. These displays are based on holographic principles[6,7,8,9,10], which have been demonstrated to encode a static three-dimensional (3D) scene with a quality indistinguishable from reality in a thin film[11] or to compress the functionality of an optical stack into a thin, lightweight holographic optical design[12,13]. Holographic displays also promise unique capabilities for near-eye displays, including per-pixel depth control, high brightness, low power and optical aberration correction capabilities, which have been explored using benchtop prototypes providing limited visual experiences[14,15,16,17,18]. Most recently, holographic near-eye displays based on thin optical waveguides have shown promise in enabling very compact form factors for near-eye displays[19,20,21], although the image quality, the ability to produce 3D colour images or the étendue achieved by these proposals have been severely limited.

A fundamental problem of all digital holographic displays is the limited space–bandwidth product, or étendue, offered by current spatial light modulators (SLMs)[22,23]. In practice, a small étendue fundamentally limits how large of a field of view and range of possible pupil positions, that is, eyebox, can be achieved simultaneously. While the field of view is crucial for providing a visually effective and immersive experience, the eyebox size is important to make this technology accessible to a diversity of users, covering a wide range of facial anatomies as well as making the visual experience robust to eye movement and device slippage on the user's head. A plethora of approaches for étendue expansion of holographic displays has been explored, including pupil replication as well as the use of static phase or amplitude masks[20,24,25,26,27,28]. By duplicating or randomly mixing the optical signal, however, these approaches do not increase the effective degrees of freedom

of the (correlated) optical signals, which is represented by the rank of the mutual intensity (MI)[29,30,31] (Supplementary Note 2). Hence, the image quality achieved by these approaches is typically poor, and perceptually important ocular parallax cues are not provided to a user[32].

One of the key challenges for achieving high image quality with holographic waveguide displays is to model the propagation of light through the system with high accuracy[20]. Non-idealities of the SLM, optical aberrations, coherence properties of the source, and many other aspects of a specific holographic display are difficult to model precisely, and minor deviations between simulated model and physical optical system severely degrade the achieved image quality. This challenge is drastically exacerbated for holographic displays using compact waveguides in large-étendue settings. A practical solution to this challenge requires a twofold approach. First, the propagation of light has to be modelled with very high accuracy. Second, such a model needs to be efficient and scalable to our large-étendue settings. Recent advances in computational optics have demonstrated that artificial intelligence (AI) methods can be used to learn accurate propagation models of coherent waves through a holographic display, substantially improving the achieved image quality[33,34]. These learned wave propagation models typically use convolutional neural networks (CNNs), trained from experimentally captured phase–intensity image pairs, to model the opto-electronic characteristics of a specific display more accurately than purely simulated models[35]. However, as we demonstrate in this Article, conventional CNN-based AI models fail to accurately predict complex light propagation in large-étendue waveguides, partly because of the incorrect assumption of the light source being fully coherent. Other important problems include the efficiency of a model, such that it can be trained within a reasonable time from a limited set of captured phase–intensity pairs and run quickly at inference time, and scalability to large-étendue settings while ensuring accuracy and efficiency.

Here, we reformulate the wave propagation learning problem as coherence retrieval based on the theory of partial coherence[31,36]. For this purpose, we derive a physics-based wave propagation model that parameterizes a low-rank approximation of the MI of the wave propagation operator inside a waveguide accounting for partial coherence, which models holographic

displays more accurately than existing coherent models. Moreover, our approach parameterizes the wave propagation through waveguides with emerging continuous implicit neural representations[37], enabling us to efficiently learn a model for partially coherent wavefront propagation at arbitrary spatial and frequency coordinates over a large étendue. Our implicit model achieves superior quality compared with existing methods; it requires an order of magnitude less training data and time than existing CNN model architectures, and its continuous nature generalizes better to unseen spatial frequencies, improving accuracy for unobserved parts of a wavefront. Along with our unique model, we design and implement a holographic display system, incorporating a holographic waveguide, holographic lens and micro-electromechanical system (MEMS) mirror. Our optical architecture provides a large effective étendue via steered illumination with an ultra-compact form factor and solves limitations of similar designs by removing unwanted diffraction noise and chromatic dispersion using a volume-holographic waveguide and holographic lens, respectively. Our architecture is inspired by synthetic aperture imaging[38], where a large synthetic aperture is formed by digitally interfering multiple smaller, mutually coherent apertures. Our goal is to form a large display eyebox—the synthetic aperture built up from multiple scanned and mutually incoherent apertures, each limited in size by the instantaneous étendue of the system. This idea is inspired by classic synthetic aperture holography[39], but we adapt this idea to modern waveguide holography systems driven by AI algorithms.

After a wave propagation model is trained in a one-time preprocessing stage, a CGH algorithm converts the target content into one or multiple phase patterns that are displayed on the SLM. Large-étendue settings require the target content to contain perceptually important visual cues that change with the pupil position, including parallax and occlusion. Traditional 3D content representations used for CGH algorithms, such as point clouds, multilayer images, polygons or Gaussians[40,41,42], however, are inadequate for this purpose. Light field representations, or holographic stereograms[43,44], meanwhile, contain the desired characteristics. Motivated by this insight, we develop a light-field-based CGH framework that uniquely models our setting where a large synthetic aperture is composed of a set of smaller, mutually incoherent apertures that are, however, partially coherent within themselves.

Our approach uniquely enables seamless, full-resolution holographic light field rendering for steered-illumination-type holographic displays.

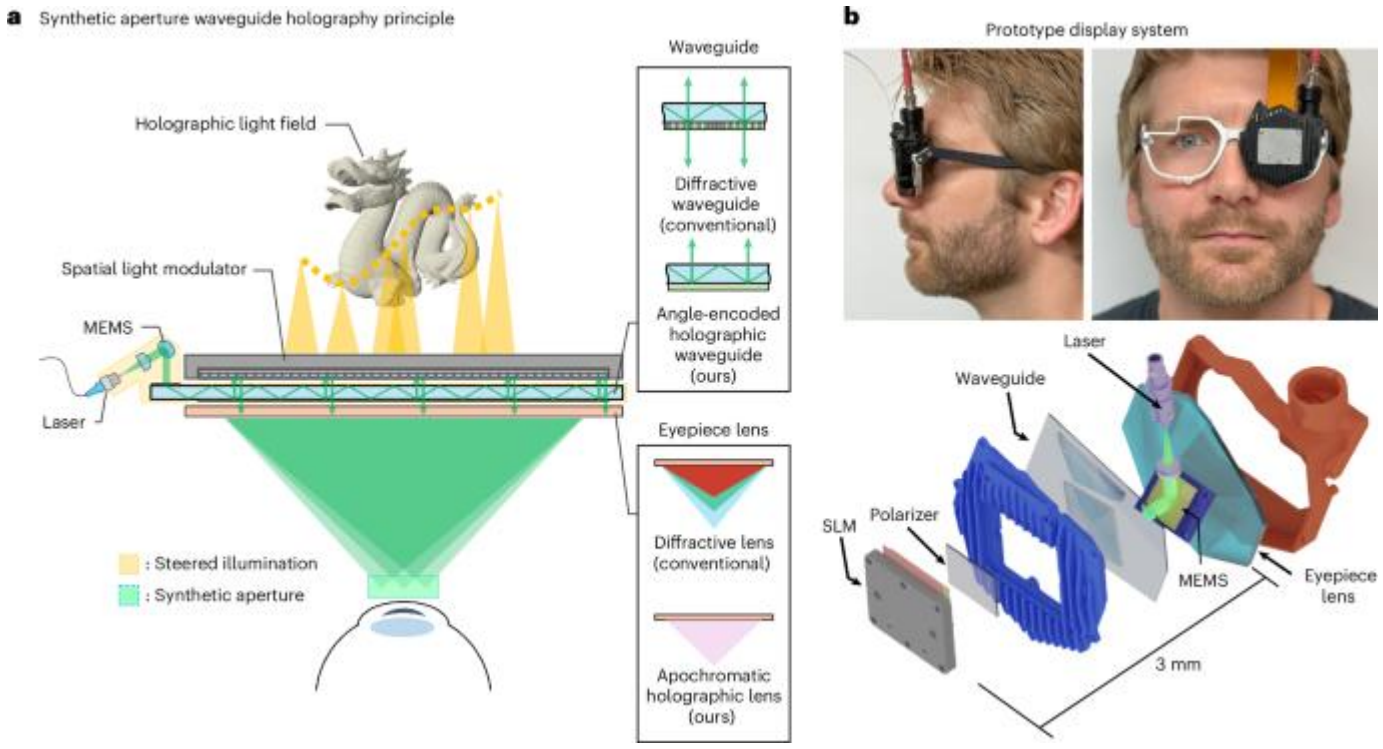
In summary, we present synthetic aperture waveguide holography as a system that combines a new and compact large-étendue waveguide architecture with AI-driven holographic algorithms. Through our prototypes, we demonstrate high 3D image quality within a 3D eyebox that is two orders of magnitude larger, marking a pivotal milestone towards practical holographic near-eye display systems.

Results

Ultra-thin full-colour 3D holographic waveguide display

Our architecture is designed to produce high-quality full-colour 3D images with large étendue in a compact device form factor to support synthetic aperture holography based on steered waveguide illumination, as shown in Fig. 1. Our waveguide can effectively increase the size of the beam without scrambling the wavefront, unlike diffusers or lens arrays[27,45]. Moreover, it allows a minimum footprint for beam steering using a MEMS mirror at the input side. For these reasons, waveguide-based steered illumination has been suggested for holographic displays[19,46]. Existing architectures, however, suffer from two problems: world-side light leakage from the waveguide and chromatic dispersion of the eyepiece lens. Here, we overcome conventional limitations with two state-of-the-art optical components to overcome conventional limitations: the angle-encoded holographic waveguide and the apochromatic holographic eyepiece lens.

Fig. 1: Holographic MR display system.



a, An illustration of the synthetic aperture waveguide holography principle. The illumination module consists of a collimated fibre-coupled laser, a MEMS mirror that steers the input light angle, and a holographic waveguide. Together, these components serve as a partially coherent backlight of the SLM. The SLM is synchronized with the MEMS mirror and creates a holographic light field, which is focused towards the user's eye using an eyepiece lens. Our design achieves an ultra-thin form factor as it consists only of flat optical elements and it does not require optical path length to form an image. The steered illumination mechanism produces a synthetic aperture, which supports a two-orders-of-magnitude larger étendue than that intrinsic to the SLM. Our angle-encoded holographic waveguide and apochromatic holographic lens design solve bidirectional diffraction noise and chromatic dispersion issues, respectively. **b**, An exploded view of the schematic image and captured photos of the holographic MR display prototype. The use of thin holographic optics achieves a total optical stack thickness of less than 3 mm (panel to lens). Top- and side-view photographs of the prototype are shown.

Our holographic waveguide shares many characteristics with conventional surface relief grating (SRG)-type waveguides. Specifically, our design uses a similar pupil replication principle and layout, which consists of an in-coupler, an exit-pupil expanding grating and an out-coupler grating. By contrast, our couplers are constructed of uniquely designed volume Bragg gratings (VBGs) instead of SRGs. SRGs used in conventional waveguides[47] have a degeneracy that supports multiple modes of diffraction. This causes a light leakage issue as the light is out-coupled bidirectionally, to both the world side and viewer's eye side. When the waveguide is used for illumination, the leakage enters the eyepiece without being modulated by the SLM, resulting in d.c. noise throughout the entire field of view. This d.c. noise significantly degrades the contrast of the displayed image[19], and it is challenging to filter out this noise as it shares the optical path with the signal. On the contrary, VBGs exhibit a type of diffraction known as Bragg diffraction[48,49], where only light that satisfies a specific incident angle and a narrow spectral bandwidth is diffracted to a single diffraction order with high efficiency. This single-direction diffraction greatly suppresses stray light and ghost images compared with conventional waveguides. In addition, we use the angle-encoded multiplexing method to cover the larger steering angle, where each grating supports only a portion of the target angular bandwidth with a specific narrow band wavelength with high efficiency. The multiplexing is performed by overlapping volume gratings with the same surface pitch but with different slant angles (see Supplementary Note 1 for additional details). This allows us to collectively cover the target angular bandwidth with high efficiency, while supporting three wavelengths of red (638 nm), green (520 nm) and blue (460 nm) without crosstalk.

Our display architecture achieves an ultra-thin form factor of only 3 mm thickness from the SLM to the eyepiece lens, including a 0.6-mm waveguide and 2-mm holographic lens. The MEMS mirror in our display steers the illumination angles incident on our SLM, which creates a synthetic aperture (that is, eyebox) of size 9 × 8 mm^2, supporting an eyebox volume that is two orders of magnitude larger than that intrinsic to the SLM. The diagonal field of view of our display is 38° (34.2° horizontal and 20.2° vertical). An in-depth discussion of the full system and additional components is found in the Methods and Supplementary Note 1.

Partially coherent implicit neural waveguide model

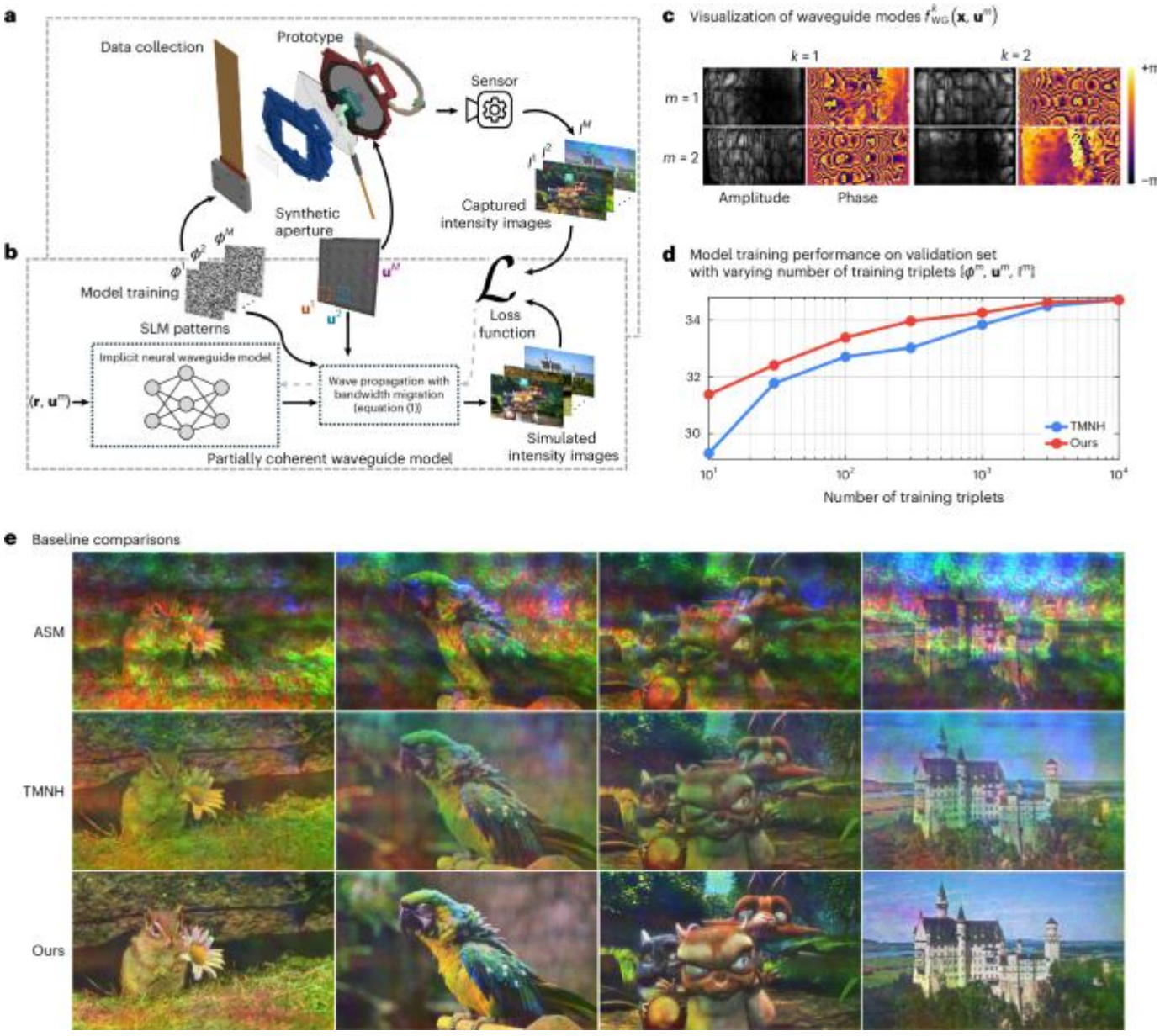
Our synthetic aperture waveguide holographic display is partially coherent because its synthetic aperture consists of a scanned set of mutually incoherent apertures. In addition, the instantaneous eyebox exhibits partial coherence due to factors such as millimetre-scale optical path length differences generated by the pupil replication process in the waveguide, mode instability of diode lasers and imperfect polarization management. Neither existing AI-based wave propagation models for coherent wavefronts[21,33,34] nor recent waveguide propagation models[20] are sufficient in adequately describing the complex behaviour of partially coherent light[50,51]. In this section, we first develop a partially coherent waveguide model to accurately characterize physical optical systems based on implicit neural representations[52]. Then, we adapt this model to our synthetic aperture waveguide holography setting. Our model can be trained automatically using camera feedback, it generalizes well to unseen spatial frequencies and it is used to produce the high-quality experimental results demonstrated in later sections.

A partially coherent wavefront can be represented by its MI or its Fourier transform—the Wigner distribution function[53,54]. With our waveguide model, we aim to model the MI of the waveguide, , where $\mathbf{r}_1$ and $\mathbf{r}_2$ are spatial coordinates on the SLM plane. This MI depends on the steered illumination angle, so it needs to be characterized for each of them separately. Representing such a high-resolution, four-dimensional function for many apertures within the synthetic aperture, however, is computationally intractable. Consider a discretized MI for 1,920 × 1,080 spatial SLM coordinates for each of 10 × 10 steering angles—the corresponding ensemble of MIs would require more than 100 terabytes of memory to be stored. Moreover, it is unclear how to smoothly interpolate these MIs between aperture positions as they are characterized at discrete positions.

To address these challenges, we introduce a low-rank implicit neural representation for the ensemble of steering-angle-dependent MIs of the waveguide. Specifically, for each steering angle $\mathbf{u}^m$, $m = 1, \ldots, M$, the MI is approximated by K coherent spatial modes that are incoherently summed[55] as , where is a single coherent mode at spatial frequency $\mathbf{u}$, $k = 1, \ldots, K$ indicates the mode index and H denotes the Hermitian transpose. Note

that coherent waveguide models are a special case of this partially coherent model, namely, those of rank 1. This low-rank MI representation is inspired by coherence retrieval methods[31]. Rather than representing the ensemble of low-rank MIs explicitly, however, we introduce a novel waveguide representation based on emerging neural implicit representations[37]. This neural-network-parameterized representation is more memory efficient than a discretized representation and it is continuous, so it can be queried at any spatial or spatial frequency (that is, aperture) coordinate. Specifically, our implicit neural representation is a multilayer perceptron (MLP) network architecture, , that represents the MI ensemble of the waveguide using network parameters Ψ. Once trained, the MLP can be queried with the input of spatial (**r**) and frequency (**u**) coordinates and it outputs the K modes of the corresponding MI. The primary benefits of this implicit neural representation over a conventional discrete representation are its memory efficiency (a few tens of megabytes versus terabytes) and, as we demonstrate in the following experiments, the fact that our representation generalizes better to aperture positions that were not part of the training data. Remarkably, even in a low-étendue setting, our implicit model achieves faster convergence and better accuracy than state-of-the-art CNN-based models based on explicit representations (Fig. 2d), so the implicit model can serve as a drop-in replacement for existing holographic displays. In a large-étendue setting, our implicit neural model is the only one achieving high-quality results.

Fig. 2: Partially coherent implicit neural waveguide model.



a, Using our prototype, we capture training and validation datasets consisting of sets of an SLM phase pattern as well as the corresponding aperture position and intensity image. The aperture positions are uniformly distributed across the synthetic aperture, enabling model training with large étendue. **b**, The captured dataset is used to train our implicit neural waveguide model. The parameters of our model are learned using backpropagation (dashed grey line) to predict the experimentally captured intensity images. **c**, A visualization of two waveguide modes of the trained model, including amplitude and phase, at two different aperture positions. Our model faithfully reconstructs

the wavefront emerging out of the waveguide, exhibiting the patch-wise wavefront shapes expected from its pupil-replicating nature. **d**, Evaluation of wave propagation models with varying training dataset sizes for a single aperture (that is, low-étendue setting). Our model achieves a better quality using a dataset size that is one magnitude lower than state-of-the-art models[34]. **e**, Experimentally captured image quality for different wave propagation models in the low-étendue setting. Our model outperforms the baselines, including the ASM[60] and the time-multiplexed neural holography model (TMNH)[34], by a large margin.

Our implicit waveguide models the MI of input light arriving at the SLM plane. Then, the phase of each coherent mode of the MI is modulated by the corresponding SLM phase pattern ϕ^m. Each of the modes continues to propagate in free space, by distance z, to the image plane in the scene. We use an off-axis angular spectrum method (ASM)[35,56] to implement this free-space wave propagation operator for each coherent mode, migrating the bandwidth over a large étendue. Finally, all propagated modes are incoherently summed over the full synthetic aperture to form the desired intensity, I_z, at distance z.

(1)

(2)

where is the coherent wave propagation operator with propagation distance z through the aperture centred at $\mathbf{u}$, and is the mean operator denoting the incoherent sum of modes. As detailed in the Methods, to optimize experimental image quality, the wave propagation operator also includes several learned components, including pupil aberrations and diffraction efficiency of the SLM.

Using these equations, we can map phase patterns shown on the SLM ϕ^m with steering angle $\mathbf{u}^m$ to the image that a user would observe. To assess our partially coherent waveguide model with respect to state-of-the-art holographic wave propagation models, we capture a dataset from our display prototype consisting of triplets, each containing an SLM phase pattern, an aperture position and the corresponding intensity at the image plane (Fig. 2a).

We then train models to predict the experimentally captured intensity images, given the known input phase patterns and aperture positions.

To evaluate and compare waveguide models in a large-étendue setting, we train ours along with baseline models on a dataset captured at 72 aperture positions with our prototype. We evaluate the generalization capabilities of these models on a test set containing nine aperture positions that were not part of the training set. Quantitative results are presented in Table 1. Previously proposed explicit CNN models generalize poorly to the unseen aperture positions. Importantly, it takes more than 2 days of training for them to converge, which is somewhat impractical. Our implicit waveguide model is smaller in size and shows better generalization capabilities than explicit models, while also converging much faster. Moreover, when used in a partially coherent configuration, the accuracy of our model is drastically improved over all coherent models, as seen by the high quality achieved on both training and test sets when used with three or six modes.

Table 1 Comparison of different waveguide and wave propagation models

Figure 2a illustrates the data collection aspect of our approach: we show a sequence of phase patterns, ϕ^m, on the SLM and capture corresponding intensity images, I^m, with a camera focused on the image plane. The implicit neural model is trained on these data, as shown in Fig. 2b. For this purpose, we simulate the forward image formation given the phase patterns ϕ^m and a random set of initialized model parameters. A loss function measures the mean-squared error between simulated and captured intensity images. The backpropagation algorithm is applied to learn all model parameters, including the partially coherent waveguide model. This model comprises a set of coherent modes , each being a complex-valued image on the SLM plane with amplitude and phase, that can be queried at the spatial frequencies corresponding to the MEMS steering angles $\mathbf{u}^m$ (Fig. 2c). In a low-étendue setting, that is, when operating with a fixed steering angle $\mathbf{u}^m$, this implicit neural model achieves a higher quality with significantly less training data compared with state-of-the-art explicit CNN models[34] (Fig. 2d). Therefore, our model serves as a drop-in replacement for existing wave propagation models

with strictly better performance. Even though the gain in peak signal-to-noise ratio for explicit and implicit models on the validation set at convergence is only a few decibel (Fig. 2d), the generalization capabilities of our implicit model are far superior to those of the explicit model. This is observed in Fig. 2e, where we show experimental results comparing a model-free free-space wave propagation operator[35], the results achieved by the coherent explicit model[34] and our partially coherent implicit model. The images used here are representative of a test set of images unseen during training and show that our model outperforms both baselines by a large margin.

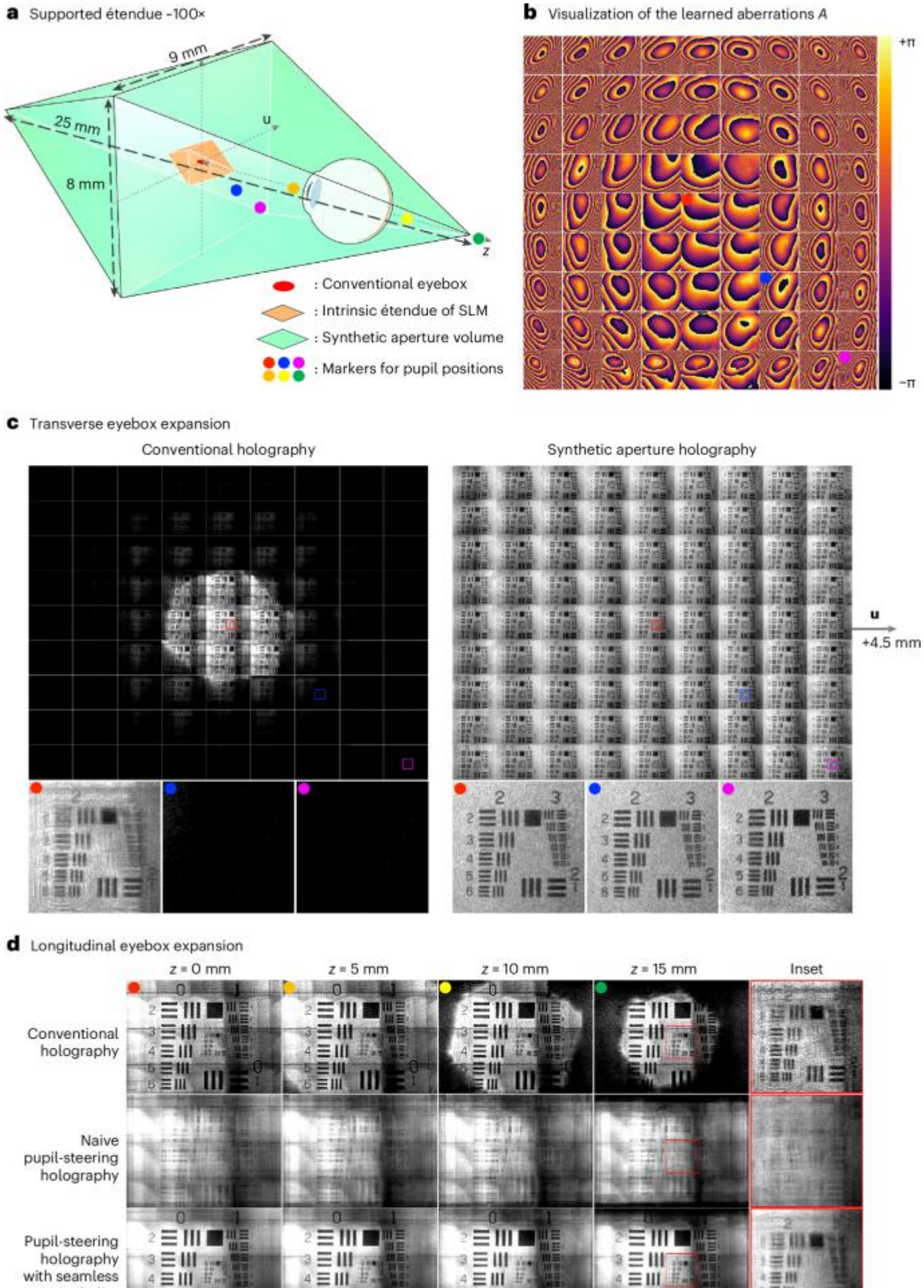
The 3D eyebox describes the volume within which a hologram is visible to the observer's eye. A large eyebox is crucial for guaranteeing high image quality when the eye moves and for making any display accessible to a diverse set of users with different binocular characteristics. Figure 3a illustrates the supported eyebox volume of our system (green volume), which is two orders of magnitude larger than that intrinsic to the SLM without steered illumination (orange volume). Moreover, our CGH framework fully utilizes the bandwidth corresponding to an eyebox size of 9 × 8 mm, unlike conventional holographic displays[16,44,57], which use only a small fraction of the available bandwidth supported by the SLM (red volume). Figure 3b visualizes images of the phase aberrations of the waveguide system learned by our model at various positions within the supported eyebox. These aberrations vary drastically over the eyebox, making it challenging to be learned using approximately shift-invariant CNN-based models. We further validate our model's ability to produce high-quality imagery across an extended étendue and compare its performance with that of conventional holography in Fig. 3c,d.

Figure 3c shows experimentally captured light fields of a two-dimensional resolution chart image located at optical infinity over the extended eyebox in the top row, while the insets in the bottom present the results at specific positions, indicated by corresponding colours. As expected, conventional holography provides a very small eyebox, which restricts the image to be seen only from the centre position. Our approach supports a significantly larger eyebox with high uniformity for transversely shifting eye positions.

Figure 3d demonstrates longitudinal eyebox expansion. Conventional holography suffers from vignetting, and naive pupil-steered holography[25] does not account for the axial shift of the eye, resulting in inconsistent image overlap. In addition, variations in the waveguide output

and aberrations across different steering states significantly degrade image quality, even when the issue of inconsistent image overlap is addressed, as shown in the third row of Fig. 3d. Our method generates robust, high-quality imagery as the eye moves along the optical axis. Note that the pupil position denoted with a blue marker is not used during model training, yet our model achieves comparable image quality and consistent interpolation of parameters, demonstrating its generalization capabilities.

Fig. 3: Experimental validation of the waveguide holography system in a large-étendue setting.


a Supported étendue ~100×
9 mm
25 mm
8 mm
u
z
: Conventional eyebox
: Intrinsic étendue of SLM
: Synthetic aperture volume
: Markers for pupil positions
b Visualization of the learned aberrations A
+π
−π
c Transverse eyebox expansion
Conventional holography
Synthetic aperture holography
u
+4.5 mm
d Longitudinal eyebox expansion
z = 0 mm
z = 5 mm
z = 10 mm
z = 15 mm
Inset
Conventional holography
Naive pupil-steering holography
Pupil-steering holography with seamless

a, Visualizations of the supported étendue (3D eyebox). Conventional holography, such as smooth phase techniques (red volume at the origin), supports only a portion of the eyebox that the SLM can produce (intrinsic étendue, orange volume). Our system supports a two-orders-of-magnitude larger eyebox than conventional holographic display systems (green volume). **b**, Visualization of our learned model across the expanded étendue. For better visualization and interpolation capabilities, refer to the Supplementary Video. **c**, Experimentally captured resolution chart images with expanded eyebox. Our system supports an eyebox size of 9 × 8 mm, which is significantly larger than the display-limited eyebox of conventional holographic displays. Using our implicit model, we can efficiently calibrate the system, achieving high image quality uniformly across the extended eyebox, as shown in the inset. **d**, Longitudinal eyebox expansion results. Markers show the corresponding pupil positions in the octahedron shown in **a**.

CGH framework for synthetic aperture waveguide holography

At runtime, a phase-retrieval-like CGH algorithm computes one or multiple phase patterns that are displayed on the SLM to create a desired intensity image, volume or light field. In the large-étendue setting, conventional 3D content representations, such as point clouds, multilayer images or polygons[40,41], are infeasible because the large synthetic aperture requires view-dependent effects to be modelled. For this reason, a light field L is a natural way to represent the target content in this setting. Existing light-field-based CGH algorithms aim to minimize the error between each ray of the target light field and the hologram converted into a light field representation[34,43,44], typically using variants of the short-time Fourier transform. Supervising the optimized SLM phase on light rays, however, is not physically meaningful because the notion of a ray does not exist in physical optics at the scale of the wavelength of visible light.

To address this issue, we propose a new light-field-based CGH framework. Unlike existing algorithms that rely on ray-based representations for light-field holograms or use random pupils in the Fourier domain[58], we supervise our

optimization reflecting the wave nature of light more accurately, by incorporating partial coherence and phase continuity at the target field. For this purpose, we formulate the physically correct partially coherent image formation of our hologram to simulate an image passing through the user's pupil, I_{holo}, and compare it with the corresponding image simulated incoherently from the light field, I_{lf}, as

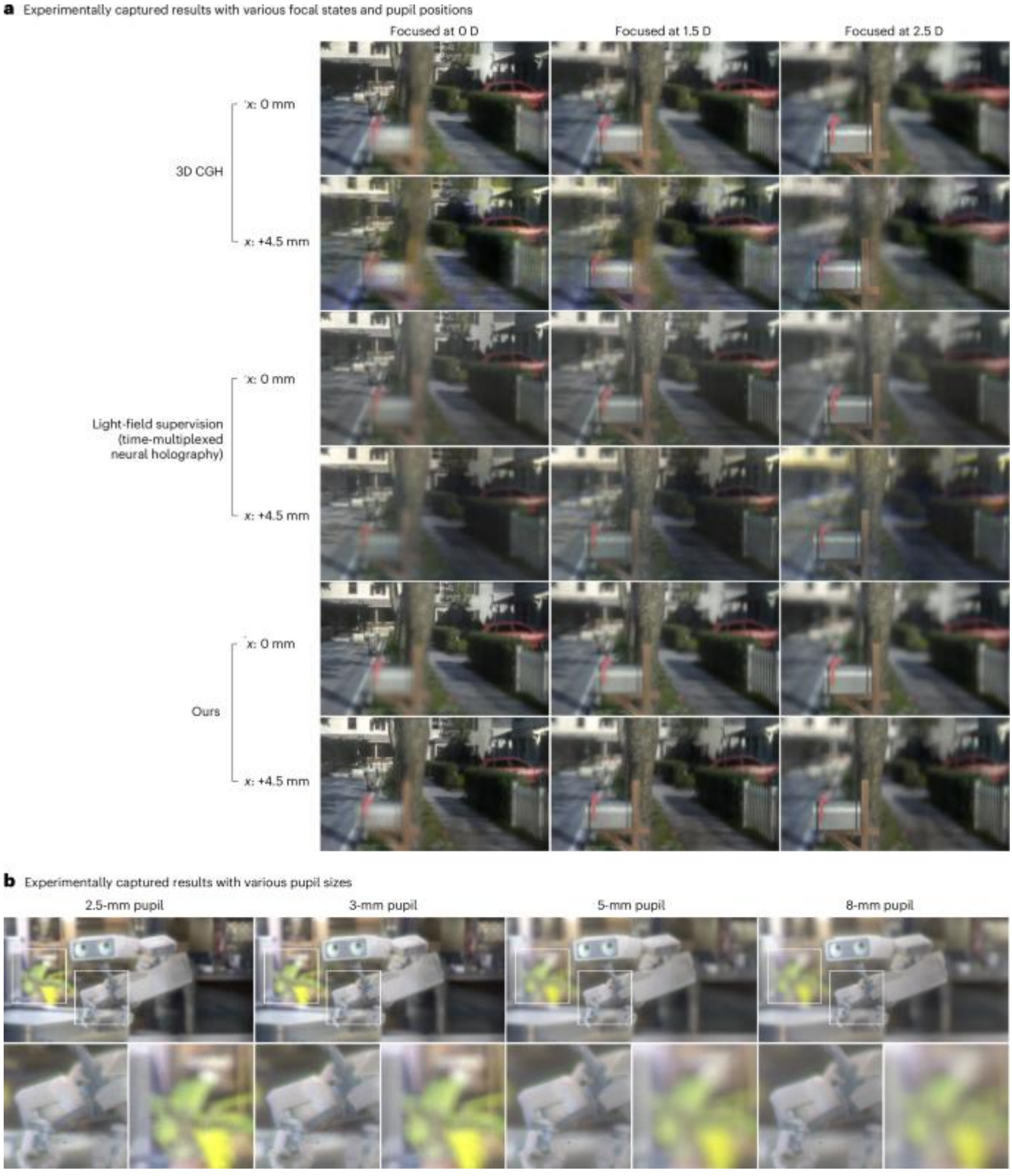
(3)
(4)
(5)

where $\mathbf{1}_{\mu}$ is the indicator function for a set of subapertures μ (a subaperture is a small aperture state spanning the pupil positions closest to the corresponding light field view), and $\mathbf{U}$ is the set of all possible subaperture combinations within our synthetic aperture. We thus aim to optimize phase patterns for any possible pupil position, diameter and shape of the user simultaneously. For this purpose, we randomly sample a batch of subaperture configurations in each iteration of our optimization routine, with the loss function measuring the mean-squared error for each subset of the apertures within the batch (see the Methods and Supplementary Note 5 for additional details).

In Fig. 4, we show experimentally captured 3D holograms. Figure 4a shows photographs captured at various focal distances—0 D (∞ m), 1.5 D (0.67 m) and 2.5 D (0.4 m)—for two different camera positions within the eyebox (that is, shifted laterally by 4.5 mm). Moreover, we compare a conventional 3D CGH algorithm[57], the state-of-the-art light-field-based CGH algorithm[34], and our CGH framework for all of these settings. Our results achieve the highest quality, exhibiting the best contrast and sharpness, and they demonstrate clear 3D refocusing capabilities as well as view-dependent parallax. In Fig. 4b, we capture a scene from a single camera position and a fixed focus with a varying camera pupil diameter. As is expected, the in-focus part of the 3D scene remains sharply focused without the conventional resolution degradation while the out-of-focus blur, that is, the depth-of-field effect, gets stronger with an increasing pupil diameter. The phase SLM patterns do not need to be recomputed for varying pupil positions, diameters or shapes, as all possible configurations are intrinsically accounted for by our unique CGH

framework. All diffraction orders created by the SLM are jointly optimized to reduce artefacts[59].

Fig. 4: Experimentally captured results in a large-étendue setting.


a Experimentally captured results with various focal states and pupil positions
Focused at 0 D
Focused at 1.5 D
Focused at 2.5 D
3D CGH
x: 0 mm
x: +4.5 mm
Light-field supervision (time-multiplexed neural holography)
x: 0 mm
x: +4.5 mm
Ours
x: 0 mm
x: +4.5 mm
b Experimentally captured results with various pupil sizes
2.5-mm pupil
3-mm pupil
5-mm pupil
8-mm pupil

a, Comparison of holographic rendering techniques using experimentally captured results with different focus states (far: 0 D, middle: 1.5 D, near: 2.5 D) and pupil positions. **b**, Experimental results with various pupil sizes demonstrate robust image quality in the focused object (left insets), with correctly represented depth of field (right insets) according to the pupil size. RGB channel images were captured separately at the same wavelength and merged to enhance the visual perception of the 3D effect and image quality (pseudo-colour).

Discussion

The co-design of our synthetic aperture waveguide holography hardware and AI-based holography algorithms enables a compact full-colour 3D MR system with large étendue, demonstrating a high-quality holographic light field across a large 3D eyebox within a compact display device form factor. Partial coherence is the key to achieving the high-quality results we demonstrate with a large étendue, because it enables the creation of light-field holograms with high-rank MI, thus offering more degrees of freedom than purely coherent image formations.

Our prototype supports a 9 × 8 mm eyebox size. In Supplementary Note 8, we derive the relationship between the transverse eyebox size e, SLM size L_{SLM}, eyepiece focal length f and the maximum steering angle θ_{scan} as

(6)

where the equality holds when the gap between the eyepiece and the SLM approaches zero (that is, zero thickness waveguide and polarizers).

Indeed, equation (6) shows that the eyebox size of our prototype approaches the theoretical upper bound we could achieve with any holographic display system, as our eyebox size is close to that of the SLM. This is particularly exciting because Fresnel holography configurations, such as ours, typically support only very small eyebox sizes.

The proposed co-design of compact large-étendue waveguide architecture and AI-driven algorithms enables compact 3D MR display modes with high image quality across a large eyebox, offering a path towards true 3D holographic MR glasses.

Methods

Experimental set-up

Our architecture is shown in Fig. 1. The total field of view of the compact prototype is 38° diagonal, with a supported eyebox size of 9 × 8 mm and an eye-relief distance range of 23–33 mm (Fig. 3). The diffraction-limited resolution of this system is 1.2 arcmin of visual angle, which is close to the human visual acuity of 20/20 vision, that is, 1 arcmin. The fibre-coupled laser is collimated using a small lens, and the MEMS mirror is tilted 45° to steer the beam directly into the waveguide.

The layout of the waveguide is similar to typical pupil-replicating waveguides[47], with three diffractive elements, an in-coupler, a pupil-expansion coupler and an out-coupler. In our compact prototype, each coupler is an angle-encoded multiplexed VBG hologram to cover the target steering angle of 20°, for three wavelengths of 460 nm, 520 nm and 638 nm. The grating is optically fabricated by recording interference pattern of two laser beams on photo-refractive polymer using custom-designed prisms[61]. The substrate thickness is 0.6 mm, and anti-reflective coating is applied for visible light. The output light is filtered using a polarizer before illuminating the SLM. The achievable frame rates of our prototype are limited by the refresh rate of the SLM (60 Hz). Yet, it is possible to achieve the full bandwidth of the MEMS mirror (400 Hz) with the use of high-frame-rate SLMs that are commercially available[62]. We do not filter higher-order copies, which contribute to the images by reducing contrast (for example, in Fig. 4, the 8-mm pupil exhibits slightly lower contrast than the 2.5-mm pupil). We note that the effect of higher diffraction orders decreases as the pixel pitch size of the SLM decreases. While our SLM has an 8-µm pixel pitch, a pixel pitch below 4.8 µm can effectively separate higher diffraction orders for pupil sizes of 3 mm.

In addition to a compact prototype, we also built a benchtop prototype as a testbed for our algorithms. The main purpose of the benchtop set-up is to test our waveguide and wave propagation models. For this prototype, we use a conventional eyepiece and SRG waveguide and expand the beam path using multiple 4-f systems to ablate the effect of undiffracted light by optically separating it from the waveguide, while keeping the specifications of wavelengths, SLM, MEMS mirror and geometric parameters similar to those of the compact prototype set-up, which uses the holographic waveguide and eyepiece. We captured Figs. 1 and 2 with our compact prototype and Figs. 3 and 4 with our benchtop prototype.

Coherent wave propagation model with bandwidth migration

In a large-étendue setting, we need to simulate a coherent wave propagation operator for a wide range of spatial frequencies. This is computationally challenging for conventional wave propagation algorithms, because a very large number of spatial positions or, similarly, spatial frequencies must be processed simultaneously. We derive a computationally efficient implementation of this wave propagation in synthetic aperture waveguide holography displays based on bandwidth migration.

The naive approach to modelling the propagation of a coherent input field f_{in} with a steering angle $\mathbf{u}^m$, for example, guided by steered illumination, is to apply a phase ramp function to the input as π. As mentioned above, however, adequately sampling this function would require a prohibitively large number of spatial frequencies to be considered. Instead, we migrate the bandwidth in the Fourier plane and use an off-axis wave propagation to sample the output field at the shifted coordinates[56]. This off-axis wave propagation with bandwidth migration can be represented as

(7)

(8)

where $\mathbf{x}$ is the spatial coordinate at the propagated plane and is the shift in the spatial domain during the propagation. Note that this calculation can be performed at the same cost as the original wave propagation, without steering. The inverse Fourier transform along the shifted frequency coordinate

does not affect the incoherent sum of each propagate mode, while avoiding aliasing in the transfer function (see Supplementary Note 4 for details). In our implementation of equations (1) and (7), we additionally parameterize and learn non-idealities of the physical optical setup, including nonlinear voltage-to-phase mapping and imperfect diffraction efficiency of the SLM. Thus, in our partially coherent propagation model, each coherent mode is propagated to the target plane at z as

(9)

where k and m are the index for coherent modes and the synthetic aperture, respectively, are the kth mode of the waveguide output and the corresponding undiffracted component from the SLM, is the nonlinear phase response of the SLM represented as a small MLP[33], and is propagation operator with learned aberrations : . These modes are added up incoherently and, optionally, an incoherent term to model additional stray light can be added as a separate mode. We note that modelling high diffraction orders and off-axis propagation are essential to capturing the accurate physics of this wave propagation, as presented in Table 1.

Implicit neural waveguide model

The waveguide output can be understood as an ensemble of beam patches shaped by multiple total internal reflections and interactions with the gratings. These interactions are governed by the diffraction efficiency of the gratings, which varies spatially owing to non-uniformities, such as fabrication-induced defects, and angularly in response to changing input angles. In addition, propagation within the waveguide necessitates bandwidth migration, as described in equation (7), which reveals predictable wavefront shifts with varying input angles. This understanding motivates our use of a learned coordinate mapping (coordinate transform MLP) that flexibly learns these shifts with angle changes, enabling scalable modelling over a large étendue.

Our model is general enough to model steered or non-steered illumination through a waveguide or for a free-space optical set-up. In all cases, we first input the spatial coordinates and steered angles into a coordinate transformation MLP with three layers, 32 features and rectified linear unit

nonlinearity. This MLP outputs transformed spatial coordinates, which are then passed through a hyperbolic tangent layer to ensure a valid range. Finally, these transformed coordinates are used to query features using multi-resolution hash encoding[37]. The features from the hash table are subsequently fed into the main waveguide MLP with 3 layers, 64 features and rectified linear unit nonlinearity, as illustrated in Fig. 2a, and outputs K complex field values at given spatial coordinates. We use $K = 4$ for the benchtop set-up and $K = 16$ for the thin benchtop set-up. As discussed in the primary text, this model significantly reduces the number of parameters that need to be optimized and, thus, provides a much more compact representation that can be optimized faster and with less training data than comparable explicit CNN-based models. Moreover, this model is continuous and naturally interpolates the low-rank MI it represents to any spatial frequency on the synthetic aperture.

We note that implicit neural representations offer a flexible design trade-off for modelling large-étendue optical systems, as memory scales with signal complexity rather than resolution. These representations have been shown to efficiently recover 3D refractive index maps and volumetric fluorescence images, outperforming classical methods in optical microscopy[63,64]. Unlike these approaches, which map 3D coordinates to a refractive index or image volume, our model maps 4D spatio-angular coordinates to the complex-valued wavefront or pupil aberrations, making it well-suited for modelling large-étendue optical systems.

Data availability

A full-colour captured dataset specific to our holographic MR display prototype and a large-étendue captured dataset are available from the corresponding authors upon request.

Code availability

Computer codes supporting the findings of this study are available online at https://github.com/choisuyeon/sawh.

Meta dishes out $250M to lure 24-year-old AI whiz kid

Mark Zuckerberg's Meta gave a 24-year-old artificial intelligence whiz a staggering $250 million compensation package, raising the bar in the recruiting wars for top talent — while also raising questions about economic inequality in an AI-dominated future.

Matt Deitke, who recently dropped out of a computer science doctoral program at the University of Washington, initially turned down Zuckerberg's "low-ball" offer of approximately $125 million over four years, according to the New York Times.

But when the Facebook founder, a former whiz kid himself, met with Deitke and doubled the offer to roughly $250 million — with potentially $100 million paid in the first year alone — the young researcher accepted what may be one of the largest employment packages in corporate history, the Times reported.

"When computer scientists are paid like professional athletes, we have reached the climax of the 'Revenge of the Nerds!'" Professor David Autor, an economist at MIT, told The Post on Friday.

Deitke's journey illustrates how quickly fortunes can be made in AI's limited talent pool.

After leaving his doctoral program, he worked at Seattle's Allen Institute for Artificial Intelligence, where he led the development of Molmo, an AI chatbot capable of processing images, sounds, and text — exactly the type of multimodal system Meta is pursuing.

In November, Deitke co-founded Vercept, a startup focused on AI agents that can autonomously perform tasks using internet-based software. With approximately 10 employees, Vercept raised $16.5 million from investors including former Google CEO Eric Schmidt.

His groundbreaking work on 3D datasets, embodied AI environments and multimodal models earned him widespread acclaim, including an Outstanding Paper Award at NeurIPS 2022. The award, one of the highest accolades in the AI

research community, is handed out to around a dozen researchers out of more than 10,000 submissions.

The deal to lock up Deitke underscores Meta's aggressive push to compete in artificial intelligence.

Meta has reportedly paid out more than $1 billion to build an all-star roster, including luring away Ruoming Pang, former head of Apple's AI models team, to join its Superintelligence Labs team with a compensation package reportedly worth more than $200 million.

The company said capital expenditures will go up to $72 billion for 2025, an increase of approximately $30 billion year-over-year, in its earnings report Wednesday.

While proponents argue that competition drives innovation, critics worry about the concentration of power among a few companies and individuals capable of shaping AI's development.

Ramesh Srinivasan, a professor of Information Studies and Design/Media Arts at UCLA and founder of the university's Digital Cultures Lab, said the direction that companies like Meta are taking with artificial intelligence is "foundational to why our economy is becoming more unequal by the day."

"These firms are awarding hundreds of millions of dollars to a handful of elite researchers while simultaneously laying off thousands of workers—many of whom, like content moderators, are not even classified as full employees," Srinivasan told the New York Post.

"These are the very jobs Meta and similar companies intend to replace with the AI systems they're aggressively developing."

Srinivasan, who advises US policymakers on technology policy and has written extensively on the societal impact of AI, said this model of development rewards those advancing large language models while "displacing and disenfranchising the workers whose labor, ironically, generated the data powering those models in the first place."

"This is cognitive task automation," he said. "It's HR, administrative work, paralegal work — even driving for Uber. If data can be collected on a job, it can

be mimicked by a machine. All of those forms of income are on the chopping block."

Asked whether universal basic income might address mass displacement, Srinivasan, who hosts the Utopias podcast, called it "highly insufficient."

"Yes, UBI gives people money, but it doesn't address the fundamental issue: no one is being paid for the data that makes these AI systems possible," he said.

On Wednesday, Zuckerberg told investors on the company's earnings call: "We're building an elite, talent-dense team. If you're going to be spending hundreds of billions of dollars on compute and building out multiple gigawatt of clusters, then it really does make sense to compete super hard and do whatever it takes to get that, you know, 50 or 70 or whatever it is, top researchers to build your team."

"There's just an absolute premium for the best and most talented people."

A Meta spokesperson referred The Post to Zuckerberg's comments to investors.

Meta's Mind Reader: Brain2Qwerty Translates Thoughts Into Text

By Luis E. Romero, Feb 19, 2025, 07:00am ESTFeb 19, 2025, 02:52pm EST

In a quiet lab in Spain, 35 volunteers spent hours typing sentences while a half-ton machine recorded their brain activity. This setup at the Basque Center on Cognition, Brain and Language gave rise to Brain2Qwerty, Meta's most ambitious neuroscience project to date. The invention is a non-invasive brain-computer interface (BCI) that decodes sentences from neural activity recorded via electroencephalography (EEG) or magnetoencephalography (MEG). Detailed in Meta's research publication, the system represents a leap forward in assistive

communication technologies, particularly for individuals with speech or motor impairments. By translating brain signals into text as users type on a QWERTY keyboard, Brain2Qwerty bridges the gap between invasive neural implants and non-invasive alternatives.

The Architecture: A Three-Stage Neural Network

Brain2Qwerty's innovation lies in its hybrid deep-learning architecture, which combines convolutional neural networks (CNNs), transformer models, and a pretrained language module. The CNN layer extracts spatial and temporal features from raw EEG/MEG data, isolating patterns linked to motor activity during typing. These signals are then processed by a transformer module to contextualize sequences—predicting words or phrases rather than isolated characters. Finally, a language model refines output by correcting errors and aligning predictions with linguistic probabilities, akin to an autocorrect for the brain. (Meta AI, 2025).

This multi-stage approach diverges from older BCI methods that relied on external stimuli or imagined movements. Instead, Brain2Qwerty leverages natural motor processes, reducing cognitive load and enabling more intuitive use. Early trials with 35 healthy participants typing memorized sentences, such as *"el procesador ejecuta la instrucción,"* demonstrated the system's ability to learn individual neural signatures for each keystroke, even correcting typographical errors mid-stream—a sign it captures both motor and cognitive intent.

The Brain's Language Blueprint: From Contextual Thought to Written Text

Brain2Qwerty's findings illuminate the hierarchical nature of language production in the brain. The researchers discovered a clear neuro-hierarchy: before producing each word, the brain first represents its context, then loads its meaning, and finally represents syllables and letters. This top-down sequence of activations precedes word production, with context representations emerging

first, followed by word-, syllable-, and letter-level representations. By chaining these elements, the brain seamlessly transforms thoughts into communication. This hierarchical process aligns with longstanding linguistic theories and provides unprecedented insight into the neural dynamics of language production.

Performance Metrics and Limitations

While Brain2Qwerty marks progress, its accuracy hinges on the imaging technology used. MEG-based decoding achieved a 32% character error rate (CER) on average, with top performers reaching 19% CER. EEG, however, lagged at 67% CER due to lower spatial resolution. For context, professional human transcribers average 8% CER, and invasive systems like Neuralink report sub-5% rates.

These results underscore MEG's superiority but also highlight challenges. Meta's current machine, built with an MEG, costs $2 million and weighs 500 kg, making it impractical for daily use. Additionally, Brain2Qwerty processes sentences after completion rather than in real time, limiting its utility for practical applications like fluid conversation. The study also excluded individuals with motor impairments, leaving open questions about its adaptability to locked-in syndrome patients or those with neurodegenerative conditions.

Ethics, Accessibility and the Path Forward

Meta emphasizes that Brain2Qwerty decodes only *intended* keystrokes, not unfiltered thoughts—a crucial distinction for privacy. Yet as BCIs evolve, ethical frameworks must address data security and consent, particularly if commercialization occurs. For now, Meta's focus remains on research, including 'transfer learning' to adapt models to new users, and collaborations to integrate Brain2Qwerty with large language models like GPT-4 or similar architectures for semantic decoding.

Hardware miniaturization is another priority. Portable MEG prototypes could democratize access, while hybrid EEG setups might balance cost and accuracy. Clinically, pairing Brain2Qwerty with eye-tracking or gesture-based systems could offer multimodal solutions for patients. As researchers note, the goal isn't to replace invasive BCIs but to expand options for those unable or unwilling to undergo surgery.

The Road Ahead

Brain2Qwerty is a milestone in non-invasive neurotechnology, yet its road to real-world impact remains long. Closing the performance gap with invasive methods, ensuring equitable access, and navigating ethical pitfalls will require interdisciplinary collaboration. But for millions awaiting communication solutions, this AI-driven interface signals hope—and a future where thoughts transcend physical limits.

NewsArtificial Intelligence

Meta Has an AI That Can Read Your Mind and Draw Your Thoughts

Meta is developing brain-scanning technology that can convert brain activity into vivid imagery in miliseconds.

By Jose Antonio Lanz
Oct 18, 2023

Image created by Decrypt using AI

Meta has unveiled a groundbreaking AI system that can almost instantaneously decode visual representations in the brain.

Meta's AI system captures thousands of brain activity measurements per second and then reconstructs how images are perceived and processed in our minds, according to a new research paper. “Overall, these results provide an important step towards the decoding—in real time—of the visual processes continuously unfolding within the human brain," the report said.

The technique leverages magnetoencephalography (MEG) to provide a real-time visual representation of thoughts.

MEG is a non-invasive neuroimaging technique that measures the magnetic fields produced by neuronal activity in the brain. By capturing these magnetic signals, MEG provides a window into brain function, allowing researchers to study and map brain activity with high temporal resolution.

The AI system consists of three main components:

- Image Encoder: This component creates a set of representations of an image, independent of the brain. It essentially breaks down the image into a format that the AI can understand and process.
- Brain Encoder: This part aligns MEG signals to the image embeddings created by the Image Encoder. It acts as a bridge, connecting the brain's activity with the image's representation.
- Image Decoder: The final component generates a plausible image based on the brain representations. It takes the processed information and reconstructs an image that mirrors the original thought.

Meta's latest innovation isn't the only recent advancement in the realm of mind-reading AI. As reported by *Decrypt*, a recent study led by the University of California at Berkeley showcased the ability of AI to recreate music by scanning brain activity. In that experiment, participants thought about Pink Floyd's "Another Brick in the Wall," and the AI was able to generate audio resembling the song using only data from the brain.

Furthermore, advancements in AI and neurotechnology have led to life-changing applications for individuals with physical disabilities. A recent report highlighted a medical team's success in implanting microchips in

a quadriplegic man's brain. Using AI, they were able to "relink" his brain to his body and spinal cord, restoring sensation and movement. Such breakthroughs hint at the transformative potential of AI in healthcare and rehabilitation.

The potential applications of such technology are vast, from enhancing virtual reality experiences to potentially aiding those who have lost their ability to speak due to brain injuries.

It's essential to approach such advancements with a balanced perspective, however. The Meta researchers noted that while the MEG decoder is swift, it's not always precise in image generation. The images it produces represent only higher-level characteristics of the perceived image, such as object categories, but might falter in detailing specifics.

The implications of this technology are profound. Beyond its immediate applications, understanding the foundations of human intelligence and developing AI systems that think like us could redefine our relationship with technology.

“The rapid advances of this technology raise several ethical considerations, and most notably, the necessity to preserve mental privacy,” the researchers warned. Ultimately, while AI can now paint our thoughts, it's up to us to ensure the canvas remains our own.

Meta’s AI achieves 80% mind-reading accuracy; What it means for the future

In partnership with the Basque Center on Cognition, Brain, and Language, Meta has created an AI model that can reconstruct sentences from brain activity with an accuracy of up to 80%.

By Waqaar Haider Feb. 17, 2025

Meta's artificial intelligence research team is advancing towards the interpretation of human thoughts. In partnership with the Basque Center on Cognition, Brain, and Language, the company has created an AI model that can reconstruct sentences from brain activity with an accuracy of up to 80%.

This research utilizes a non-invasive method for recording brain activity and, as stated by the company, may lead to technologies that assist individuals who have lost their ability to speak.

Unlike current brain-computer interfaces that typically necessitate invasive procedures, Meta's method employs magnetoencephalography (MEG) and electroencephalography (EEG), which allow for the measurement of brain activity without surgical intervention.

The AI model was trained using brain recordings from 35 participants as they typed sentences. When evaluated with new sentences, Meta asserts that it can accurately predict up to 80% of the characters typed using MEG data, demonstrating at least double the effectiveness compared to EEG-based decoding.

However, this method does have its constraints; MEG requires a magnetically shielded environment, and participants must remain still to ensure precise readings. Additionally, the technology has only been tested on healthy subjects, leaving its efficacy for individuals with brain injuries uncertain.

Beyond the translation of thoughts into text, Meta's AI is also aiding researchers in understanding the process by which the brain converts ideas into language. The AI model scrutinizes MEG recordings, monitoring brain activity at a millisecond resolution.

It uncovers how the brain changes abstract thoughts into words, syllables, and even specific finger movements during typing. A significant discovery is that the brain employs a 'dynamic neural code'—a system that links various stages of language development while retaining access to previous information.

This may elucidate how individuals effortlessly construct sentences while speaking or typing. Meta's research reinforces the potential for AI to eventually facilitate non-invasive brain-computer interfaces for those unable to communicate verbally. However, the technology is not yet prepared for practical application.

Research Toward a real-time decoding of images from brain activity

October 18, 2023•

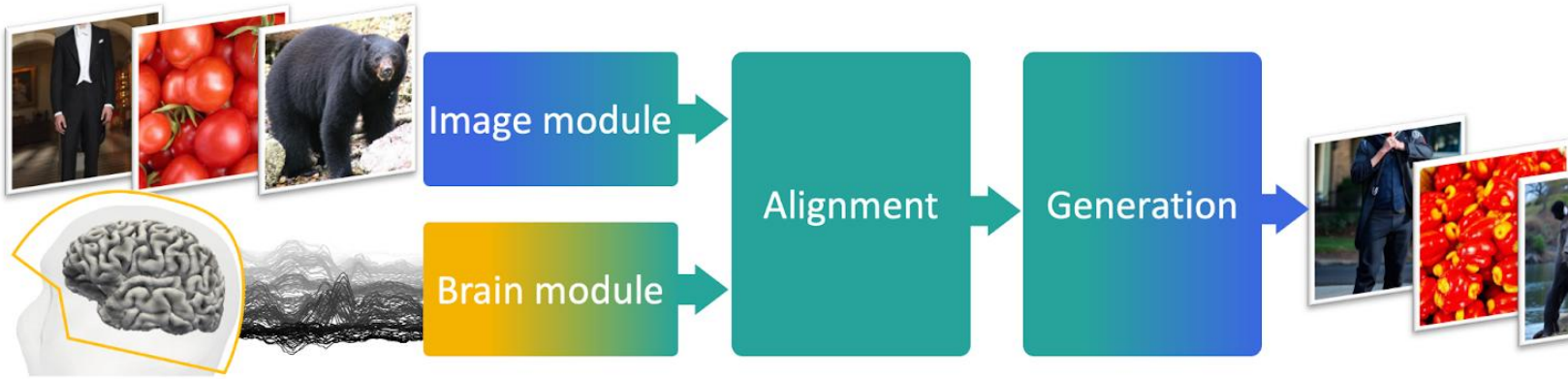
At every moment of every day, our brains meticulously sculpt a wealth of sensory signals into meaningful representations of the world around us. Yet how this continuous process actually works remains poorly understood.

Today, Meta is announcing an important milestone in the pursuit of that fundamental question. Using magnetoencephalography (MEG), a non-invasive neuroimaging technique in which thousands of brain activity measurements are taken per second, we showcase an AI system capable of decoding the unfolding of visual representations in the brain with an unprecedented temporal resolution.

This AI system can be deployed in real time to reconstruct, from brain activity, the images perceived and processed by the brain at each instant. This opens up an important avenue to help the scientific community understand how images are represented in the brain, and

then used as foundations of human intelligence. Longer term, it may also provide a stepping stone toward non-invasive brain-computer interfaces in a clinical setting that could help people who, after suffering a brain lesion, have lost their ability to speak.

Leveraging our recent architecture trained to decode speech perception from MEG signals, we develop a three-part system consisting of an image encoder, a brain encoder, and an image decoder. The image encoder builds a rich set of representations of the image independently of the brain. The brain encoder then learns to align MEG signals to these image embeddings. Finally, the image decoder generates a plausible image given these brain representations.



MEG recordings are continuously aligned to the deep representation of the images, which can then condition the generation of images at each instant.

We train this architecture on a public dataset of MEG recordings acquired from healthy volunteers and released by Things, an international consortium of academic researchers sharing experimental data based on the same image database.

We first compare the decoding performance obtained with a variety of pretrained image modules and show that the brain signals best align with modern computer vision AI systems like DINOv2, a recent self-supervised architecture able to learn rich visual representations without any human annotations. This result confirms that self-supervised learning leads AI systems to learn brain-like representations: The artificial neurons in the algorithm tend to be activated similarly to the physical neurons of the brain in response to the same image.

The images that volunteer participants see (left) and those decoded from MEG activity at each instant of time (right). Each image is presented approximately every 1.5 seconds.

This functional alignment between such AI systems and the brain can then be used to guide the generation of an image similar to what the participants see in the scanner. While our results show that images are better decoded with functional Magnetic Resonance Imaging (fMRI), our MEG decoder can be used at every instant of time and thus produces a continuous flux of images decoded from brain activity.

The images that volunteer participants see (left) and those decoded from fMRI activity (right).

While the generated images remain imperfect, the results suggest that the reconstructed image preserves a rich set of high-level features, such as object categories. However, the AI system often generates inaccurate low-level features by misplacing or mis-orienting some objects in the generated images. In particular, using the Natural Scene Dataset, we show that images generated from MEG decoding remain less precise than the decoding obtained with fMRI, a comparably slow-paced but spatially precise neuroimaging technique.

Overall, our results show that MEG can be used to decipher, with millisecond precision, the rise of complex representations generated in the brain. More generally, this research strengthens Meta's long-term research initiative to understand the foundations of human intelligence, identify its similarities as well as differences compared to current machine learning algorithms, and ultimately guide the development of AI systems designed to learn and reason like humans.

Research

Using AI to decode speech from brain activity

August 31, 2022

Every year, more than 69 million people around the world suffer traumatic brain injury, which leaves many of them unable to communicate through speech, typing, or gestures. These people's lives could dramatically improve if researchers developed a technology to decode language directly from noninvasive brain recordings. Today, we're sharing research that takes a step toward this goal. We've developed an AI model that can decode speech from noninvasive recordings of brain activity.

From three seconds of brain activity, our results show that our model can decode the corresponding speech segments with up to 73 percent top-10 accuracy from a vocabulary of 793 words, i.e., a large portion of the words we typically use on a day-to-day basis.

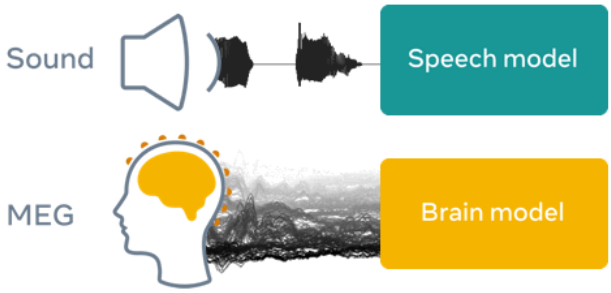
Decoding speech from brain activity has been a long-standing goal of neuroscientists and clinicians, but most of the progress has relied on invasive brain-recording techniques, such as stereotactic electroencephalography and electrocorticography. These devices provide clearer signals than noninvasive methods but require neurosurgical interventions. While results from that work suggest that decoding speech from recordings of brain activity is feasible, decoding speech with noninvasive approaches would provide a safer, more scalable solution that could ultimately benefit many more people. This is very challenging, however, since noninvasive recordings are notoriously noisy and can greatly vary across recording sessions and individuals for a variety of reasons, including differences in each person's brain and where the sensors are placed.

In our work, we address these challenges by creating a deep learning model trained with contrastive learning and then use it to maximally align noninvasive brain recordings and speech sounds. To do this, we use wave2vec 2.0, an open source, self-supervised learning model developed by our FAIR team in 2020. We then use this model to identify the complex representations of speech in the brains of volunteers listening to audiobooks.

We focused on two noninvasive technologies: electroencephalography and magnetoencephalography (EEG and MEG, for short), which measure the fluctuations of electric and magnetic fields elicited by neuronal activity, respectively. In practice, both systems can take approximately 1,000 snapshots of macroscopic brain activity every second, using hundreds of sensors.

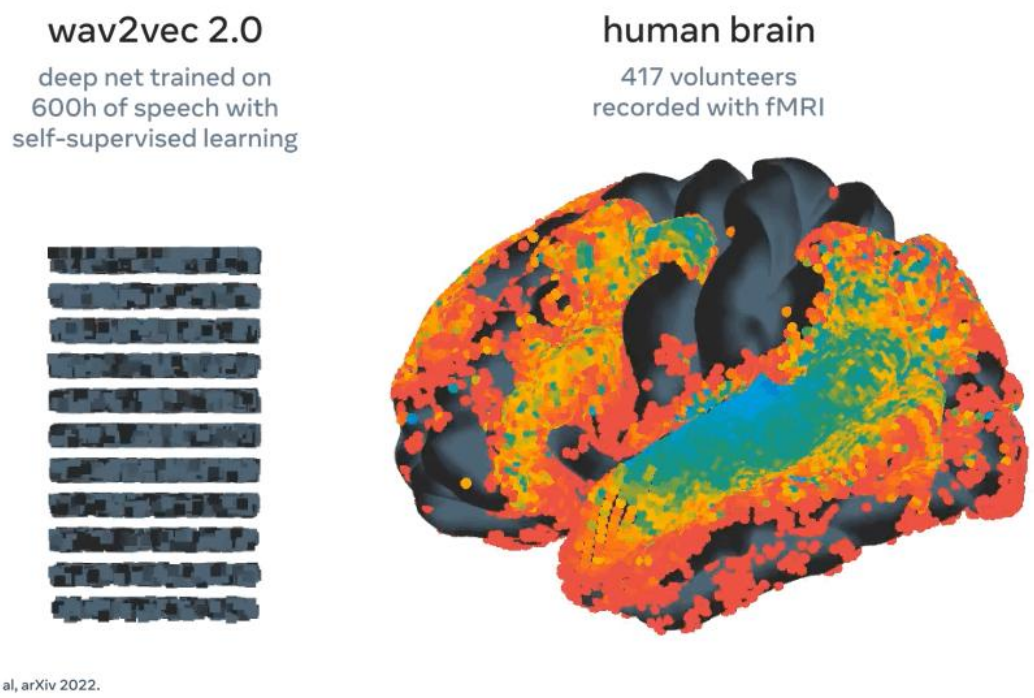
We leveraged four open source EEG and MEG datasets from academic institutions, capitalizing on more than 150 hours of recordings of 169 healthy volunteers listening to audiobooks and isolated sentences in English and Dutch.

We then input those EEG and MEG recordings into a “brain” model, which consists of a standard deep convolutional network with residual connections. EEG and MEG recordings are known to vary extensively across individuals because of individual brain anatomy, differences in the location and timing of neural functions across brain regions, and the position of the sensors during a recording session. In practice, this means that analyzing brain data generally requires a complex engineering pipeline crafted to realign brain signals on a template brain. In previous studies, brain decoders were trained on a small number of recordings to predict a limited set of speech features, such as part-of-speech categories or words from a small vocabulary. For our research, we designed a new subject-embedding layer, which is trained end-to-end to align all brain recordings in a common space.



To decode speech from noninvasive brain signals, we train a model with contrastive learning to align speech and its corresponding brain activity.

Finally, our architecture learns to align the output of this brain model to the deep representations of the speech sounds that were presented to the participants. In our previous work, we used wav2vec 2.0 to show that this speech algorithm automatically learns to generate representations of speech that align with those of the brain. The emergence of “brainlike” representations of speech in wav2vec 2.0 made it a natural choice to build our decoder, because it helps to know which representations we should try to extract from brain signals.



We recently showed that the activations of wav2vec 2.0 (left) map onto the brain (right) in response to the same speech sounds. The representations of the first layers of this algorithm (cool colors) map onto the early auditory cortex, whereas the deepest layers map onto high-level brain regions (e.g. prefrontal and parietal cortex).

After training, our system performs what's known as zero-shot classification: Given a snippet of brain activity, it can determine from a large pool of new audio clips which one the person actually heard. From there, the algorithm infers the words the person has most likely heard. This is an exciting step because it shows AI can successfully learn to decode noisy and variable noninvasive recordings of brain activity when speech is perceived. The next step is to see whether we can extend this model to directly decode speech from brain activity without needing the pool of audio clips, i.e., to move toward a safe and versatile speech decoder.

Our analyses further show that several components of our algorithm, including the use of wav2vec 2.0 and the subject layer, were beneficial to decoding performance. Furthermore, we show that our algorithm improves with the number of EEG and MEG recordings. Practically, this means that our approach benefits from the pulling of large amounts of heterogeneous data, and could, in principle, help improve the decoding of small datasets. This is important because, in many cases, it can be hard to collect a lot of data for a given participant. For example, it isn't practical to require patients to spend dozens of hours in a scanner to check whether the system works for them. Instead, algorithms could be pretrained on large datasets inclusive of many individuals and conditions, and then support the decoding of brain activity for a new patient with little data.

Using AI to understand how the brain works

The results of our research are encouraging because they show that self-supervised trained AI can successfully decode perceived speech from noninvasive recordings of brain activity, despite the noise and variability inherent in those data. These results are only a first step, however. In this work, we focused on decoding speech perception, but the ultimate goal of enabling patient communication will require extending this work to speech production. This line of research could even reach beyond assisting patients to potentially include enabling new ways of interacting with computers.

More generally, our work is a part of the broader effort by the scientific community to use AI to better understand the human brain. We're sharing this research openly to accelerate progress on the challenges still ahead. We look forward to working together and contributing to the research community in this area.

Learn more by reading our paper about decoding speech from noninvasive brain recordings.

We'd like to acknowledge contributions to this research from Alexandre Défossez, Charlotte Caucheteux, Ori Kabeli, and Jérémy Rapin.

Data were provided (in part) by the Donders Institute for Brain, Cognition and Behaviour; New York University; the University of Michigan; Trinity College Dublin; and the University of Rochester

Research

Studying the brain to build AI that processes language as people do

April 28th, 2022

AI has made impressive strides in recent years, but it's still far from learning language as efficiently as humans. For instance, children learn that "orange" can refer to both a fruit and color from a few examples, but modern AI systems can't do this nearly as efficiently as people. This has led many researchers to wonder: Can studying the human brain help to build AI systems that can learn and reason like people do?

Today, Meta AI is announcing a long-term research initiative to better understand how the human brain processes language. In collaboration with neuroimaging center Neurospin (CEA) and INRIA we're comparing how AI language models and the brain respond to the same spoken or written sentences. We'll use insights from this work to guide the development of AI that processes speech and text as efficiently as

people. Over the past two years, we've applied deep learning techniques to public neuroimaging data sets to analyze how the brain processes words and sentences.

The data sets were collected and shared by several academic institutions, including Max Planck Institute for Psycholinguistics and Princeton University. Each institution collected and shared the data sets with informed consent from the volunteers in accordance with legal policies as approved by their respective ethical committees, including the consent obtained from the study participants.

Our comparison between brains and language models have already led to valuable insights:

- Language models that most closely resemble brain activity are those that best predict the next word from context (e.g. once upon a …time). Prediction based on partially observable inputs is at the core of self-supervised learning (SSL) in AI and may be key to how people learn language.

- However, we discover that specific regions in the brain anticipates words and ideas far ahead in time, while most language models today are typically trained to predict the very next word. Unlocking this long-range forecasting capability could help improve modern AI language models.

Of course, we're only scratching the surface — there's still a lot we don't understand about how the brain functions, and our research is ongoing. Now, our collaborators at NeuroSpin are creating an original neuroimaging data set to expand this research. We'll be open-sourcing the data set, deep learning models, code, and research papers resulting from this effort to help spur discoveries in both AI and neuroscience communities. All of this work is part of Meta AI's broader investments toward human-level AI that learns from limited to no supervision

> “Understanding the origins of human intelligence is the grand challenge for science in the 21st century. It is very exciting to see artificial neural networks for language getting closer and closer to mimicking the activity of the human brain, and thereby shedding new light on how thought might be implemented in neural tissue.”
>
> Stanislas Dehaene, Professor at Collège de France and Director of NeuroSpin

Using deep learning to analyze complex brain signals

Our work is a part of the broader effort by the scientific community to use AI to better understand the brain. Neuroscientists have historically faced major limitations in analyzing brain signals — let alone comparing them with AI models. Studying neuronal activity and brain imaging is a time- and resource-intensive process, requiring heavy machinery to analyze neuronal activity, which is often opaque and noisy. Designing language experiments to measure brain responses in a controlled way can be painstaking too. For example, in classical language studies, sentences must match in complexity, and words must match frequency or number of letters, to allow a meaningful comparison of brain responses.

The rise of deep learning, where multiple layers of neural networks work together to learn, is rapidly alleviating these issues. This approach highlights where and when perceptual representations of words and sentences are generated in the brain when a volunteer reads or listens to a story.

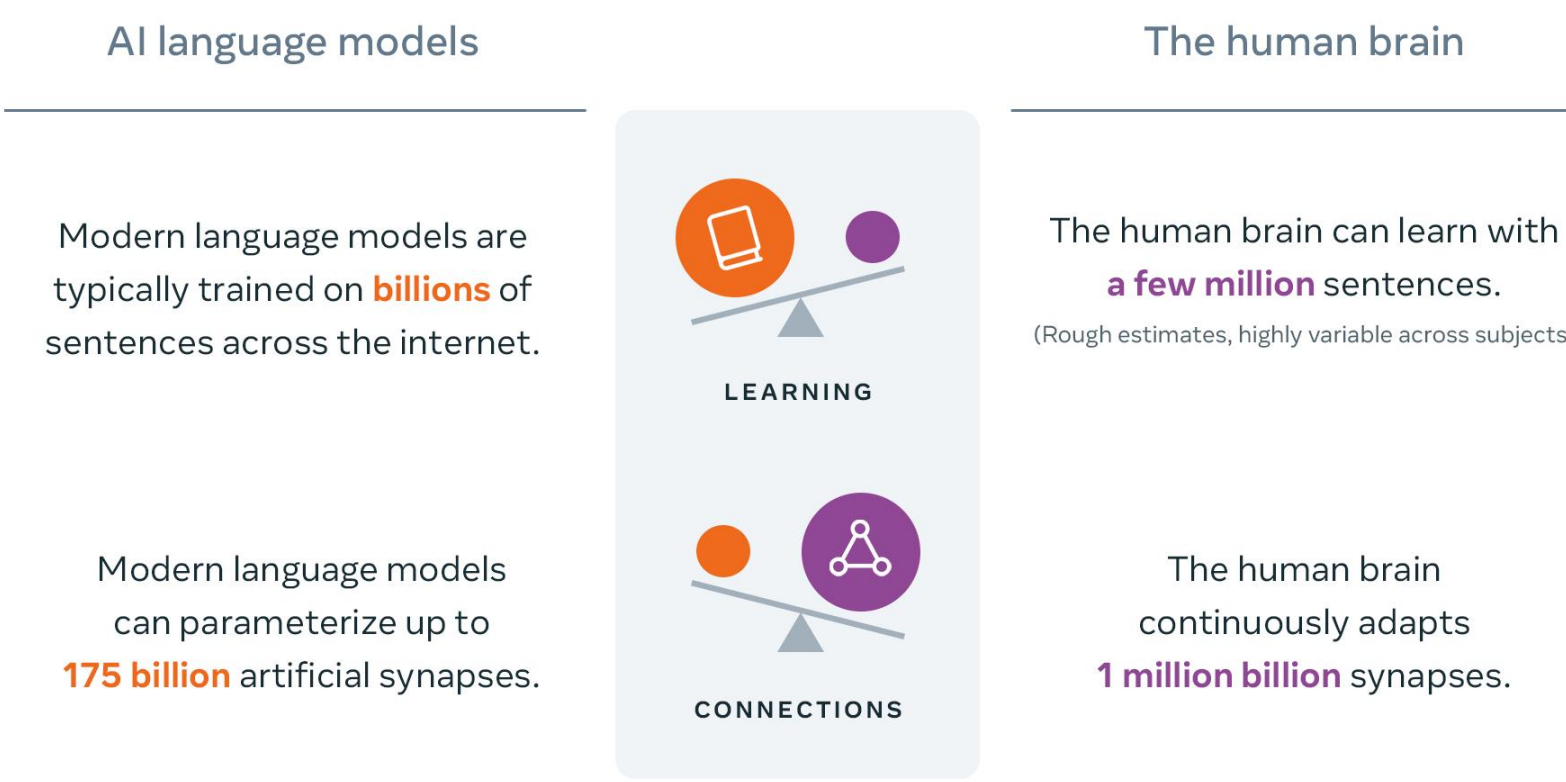
With magnetoencephalography, we can identify the brain responses to individual words and sentences at every millisecond. We can then compare areas in the brain to modern language algorithms.

Deep learning systems require a lot of data to ensure accuracy. Functional magnetic resonance imaging (fMRI) studies capture only a few snapshots of brain activities, typically from a small sample size. To meet the demanding quantity of data required for deep learning, our team not only models thousands of brain scans recorded from public data sets using fMRI but also simultaneously models them using magnetoencephalography (MEG), a scanner that takes snapshots of brain activity every millisecond — faster than a blink of an eye. In combination, these neuroimaging devices provide the large neuroimaging data necessary to detect where and in what order the activations take place in the brain. This is key to parsing the algorithm of human cognition.

In several studies, we've discovered the brain is systematically organized in a hierarchy that's strikingly similar to AI language models (here, here, and here). For example, linguists have long predicted that language processing is characterized by a sequence of sensory and lexical computations, before words can be combined into meaningful sentences. Our comparison between deep language models and the brain precisely validates this sequence of computations. When reading a word, the brain first produces representations that are similar to deep convolutional networks trained to recognize characters in the early visual cortices. These brain activations are then transformed along the visual hierarchy into lexical representations akin to word embeddings. Finally, a distributed cortical network generates neural representations that correlates with the middle and final layers of deep language models. Deep learning tools have made it possible to clarify the hierarchy of the brain in ways that wasn't possible before.

Predicting far beyond the next word

A systematic comparison between dozens of deep language models shows that the better they predict words from context, the more their representations correlate with the brain. We found this after analyzing brain activations of here, 200 volunteers in a simple reading task. A similar discovery was independently made by a team at MIT a week apart from ours, further validating this exciting direction. These similar studies provide reassurance that the AI community is on the right path with using SSL toward human-level AI.



But finding similarities isn't enough to grasp the principles of language understanding. The computational differences between biological and artificial neural networks are key to improving existing models and building new more intelligent language models. Recently, we revealed evidence of long-range predictions in the brain, which is an ability that still challenges today's language models. For instance, consider the phrase "Once upon a …" Most language models today would typically predict the next word, "time," but their ability to anticipate complex ideas, plots, and narratives like people do is still limited.

To explore this issue, together with INRIA, we compared a variety of language models to the brain responses of 345 volunteers, who listened to complex narratives while being recorded with fMRI. We enhance those models with long-range predictions to track forecasts in the brain.

Most language models today would typically predict the next word, "time," but their ability to anticipate complex ideas, plots, and narratives like people is still limited. This is something we believe the human brain is automatically able to do – not only predict the next word or sound but also the next idea.

Our results show that specific brain regions, such as the prefrontal and parietal cortices, are best accounted for by language models enhanced with deep representations of far-off words in the future. These results shed light on the computational organization of the

human brain and its inherently predictive nature and pave the way toward iimproving current AI models.

Toward human-level AI

Overall, these studies support an exciting possibility — there are, in fact, quantifiable similarities between brains and AI models. And these similarities can help generate new insights about how the brain functions. This opens new avenues, where neuroscience will guide the development of more intelligent AI, and where, in turn, AI will help uncover the wonders of the brain.