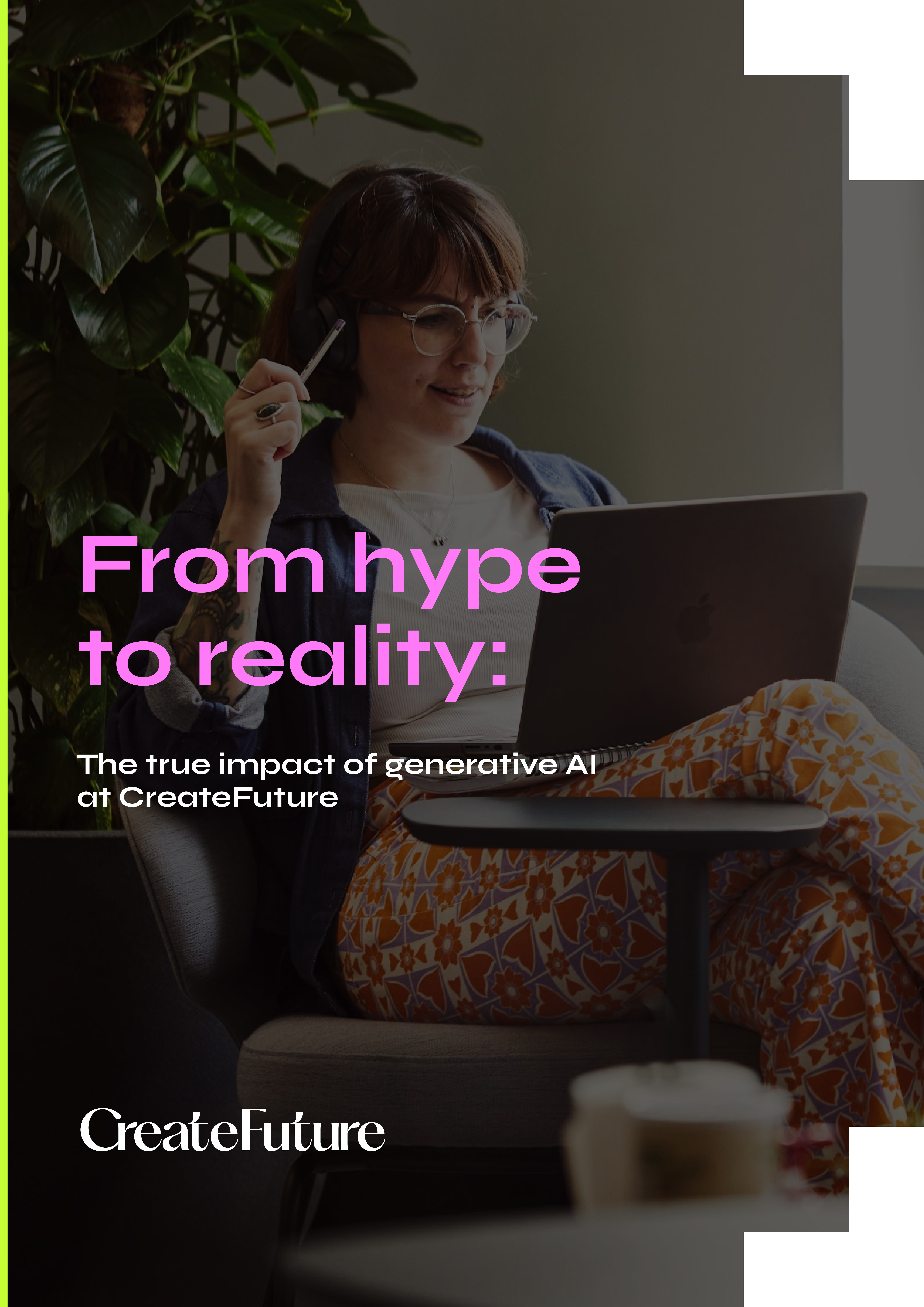




From hype to reality:

The true impact of generative AI
at CreateFuture

CreateFuture

The true impact of generative AI at CreateFuture

At the beginning of 2024, we had some opinions about AI.

Much like the rest of the world, we'd been closely watching, learning from, and trialling our own approaches to AI adoption since it came to even greater prominence in November 2022 with the launch of Open AI's ChatGPT. Since then we've been curious about studies showing its increasing importance – including one that said 92% of US-based developers were already using AI coding tools both in and outside of work.

We saw major players like GitHub, Microsoft, Atlassian and Google creating their own AI tools. Or using generative AI to add functionality to existing tools for software development.

We noticed some of our people beginning to experiment with generative AI – and others wondering aloud if it was just a flash in the pan.

But the more we saw our people – and some of our partners and competitors – experimenting with AI, the more we began to develop some new hypotheses.

Our three AI hypotheses:

1. Generative AI can significantly speed up key product engineering tasks.
2. Generative AI can help our people complete key product engineering tasks to a higher quality.
3. All the AI tools in the world are useless if you haven't got the right culture, training, L&D, approaches and mindset in place.

Too many assumptions. Too few facts.

We don't sell hype. If we're going to help our clients implement generative AI, we need to know that it can consistently deliver on its promises.

While we were seeing more and more research into the role AI can play in business operations, it was still difficult to find unbiased/reliable studies - many of these feeling like sales material - that covered the scope we needed. Studies that could help us create a methodology that would maximise the value of AI for our clients – or even for ourselves.

As a result, we needed to create our own research. Conduct a scientific study that would allow us to understand how AI could benefit our processes at CreateFuture. We also wanted to be able to prove or disprove our hunches about the tech in a realistic setting - i.e. within a fully operational business that faces a host of opportunities and challenges on a daily basis. To test the hypotheses we had created and help us refine them.

The results of that study were groundbreaking for us. And they're already shaping the way we approach AI implementation for our clients.

THE STUDY

Could our people prove our hypotheses?

Our goal was to create a study that would prove our hypotheses, set the course for AI adoption at CreateFuture and shape future strategies for our clients.

Which meant we needed to develop an airtight methodology.

Why not just run a pilot programme?

Approaching this as a scientific study had three major benefits.

✓ Justifying the spend

The cost of AI is likely to fall significantly over the next few years as providers compete to win market share. But for now, implementing AI on an organisation-wide scale can still be a significant expense.

To justify that spend, we needed hard figures. Not vague hype.

✓ Navigating the trough of disillusionment

On the Gartner Hype Cycle, AI is currently coasting through the Trough of Disillusionment – the point where some early implementations fail. It's at the point when the market begins to understand that many of the early claims about AI's ability were overblown or inaccurate.

At this point in the Hype Cycle, we need objectivity more than ever. If we want to navigate the Trough of Disillusionment and reach the Plateau of Enlightenment, we need a clear understanding of what AI can and can't do – untouched by either positive or negative hype.

✓ Building excitement and getting employee buy-in

AI implementation is a scary prospect for businesses and employees alike.

Employees rarely enjoy having a new technology forced upon them. Which means they tend to be highly wary of claims that don't seem realistic.

If we were going to ask our people to come on this journey with us, we needed to show them that it would be worth it. We needed to show that AI was going to make their working lives measurably better.

The study design

We split our study into two stages.

Stage one:

Going in cold

Thirty CreateFuture team members - product managers, front-end web software engineers, and back-end Java engineers with little to no prior experience using AI - were asked to complete a series of tasks. Some tasks were performed using the tech and some were performed without it.

Stage two:

Education and preparation

Based on the results of the first stage, we repeated the experiment with a further 23 software engineers – only this time we gave them some training on using generative AI tools.



The problems

During both stages of the experiment, we asked each participant to complete tasks from six different problem areas:

1. AI-assisted user story and acceptance

criteria writing: Participants created a series of user stories and acceptance criteria. We measured the quality of their performance based on accuracy, clarity and adherence to predefined standards.

2. Unit test writing: Participants wrote unit tests for sample projects, and we measured their performance based on the quality of the tests and the time taken to achieve specific coverage metrics.

3. Code conversion: Participants converted code from one programming language to another. Their performance was measured based on speed and accuracy.

4. Feature build in existing codebase:

Participants were asked to develop new features within an existing codebase. Their performance was measured based on how well the new features integrated and functioned.

5. Modernising code: Participants refactored a poorly written codebase. The focus was on the readability, maintainability and performance of the refactored code.

6. Epic review: The final experiment assessed the efficiency and thoroughness of using AI tools in reviewing comprehensive epics, which are large, high-level requirement documents, which are broken down into multiple user stories. These documents are collated to steer feature functionality and task prioritisation.

For each problem area, we asked each of the participants to complete two tasks – Task A and Task B.

To reduce bias, we made sure to alternate which tasks were completed with AI. In other words, some users would complete Task A without AI and Task B with AI, while others would complete Task A with AI and Task B without AI.

The evaluation

We measured each participant's performance in two ways:



Time

How long did it take to complete the task?



Quality

How did an independent evaluator rate the quality of the completed task, based on a blind reading?

We also dove into the qualitative results a little further using a survey, which measured participants' attitudes to AI before and their suggestions for improving AI adoption in the future.

What did we learn about generative AI?

Lesson one:

Using AI speeds up key product engineering tasks.

The starting hypothesis:

Generative AI can speed up key product engineering tasks

If there's one thing product engineering teams lack, it's time.

Products that launch on schedule carve out new niches in cluttered markets and win over customers. Products that launch late often fall into the shadow of the competition – or, worse still, disappear without a trace.

Which means that, at CreateFuture, we're always looking for ways to work more efficiently.

We've found ways to optimise this in the past – carefully crafted methodologies, efficient processes and extra-talented specialists are all crucial. But there's always room for more.

We were intrigued to see whether AI could make an impact.

Expectations were high: Accenture once estimated that as much as 40% of all working hours would be augmented or automated by generative AI.

Those claims become even more impressive once you focus specifically on software engineering: One McKinsey study claimed that developers can complete coding tasks up to twice as fast with generative AI.

Of course, the aim wasn't just to see how generative AI affected life for devs. We wanted to see if AI could enhance the wider product development process for other job roles too.

On the surface, it's simple:
Using AI tools reduces the time it takes to complete key product engineering tasks.

Task	Reduction in time taken to complete task
Unit test writing	14.46%
Modernising code	39.08%
Feature build	18.86%
Code conversion	41.76%

But the reality is more complex...

Examining the results began to reveal a crucial distinction – one that may help us understand why generative AI adoption might fail. And, of course, how to make sure it succeeds in the future.

“Intuitive” AI users save time. Slower learners don’t.

Overall, using generative AI helped participants complete the tasks faster.

But the figures above hide a crucial finding: a huge gap between the results of those who struggled to use the AI tools effectively, and those who found them relatively easy to use.

Our study showed that participants who managed to get to grips with AI quickly – we’ll call them “intuitive” AI users – made significant time savings.

But the participants who struggled to use their AI tools effectively saw little in the way of time savings.

Let’s look at what happens when we only look at the participants who had better AI scores, removing those who struggled to use their AI tools.

Task	Reduction in time taken to complete task	Reduction in time taken to complete task by participants with better AI scores
Unit test writing	14.46%	35.29%
Modernising code	39.08%	64.71%
Feature build	18.86%	47.06%
Code conversion	41.76%	58.82%

The change is staggering.

Our “intuitive” AI users managed to complete their tasks in half the time – sometimes even less – when using AI.

In some ways, this might seem like an obvious outcome: Someone who is good at using a tool to its full potential will likely see greater benefits from that tool than someone who struggles to use it.

But it begs the question: How many generative AI implementations are failing – or simply underperforming – because users don’t know how to use their AI tools properly?

How many generative AI implementations are failing – or simply underperforming – because users don’t know how to use their AI tools properly?

Could AI training close the gap?

As we entered the second round of testing, we had one question in mind.

Would AI become more powerful if we gave the cohort basic training in generative AI before they began tackling the tasks?

The results were clear:

In general, AI training magnifies generative AI time savings

Task	Time reduction for participants who received generative AI training
Unit test writing	39.61%
Modernising code	31.84%
Feature build	39.85%
Code conversion	55.64%

We’re still uncertain why training seemed to decrease the effects of generative AI use when modernising code – this will be an area that requires further study in the future. However our initial conjecture - which will need more testing - is that the presence of humans may add a degree of variance, which may be amplified when using a tech like AI.

But in every other area the results are clear: Even a small amount of training in using generative AI helps users complete key product engineering tasks faster.

And training doesn’t just improve scores for those who struggled to use AI. Even those who took to AI tools relatively intuitively benefited from this extra support.

Task	Reduction in time taken to complete task by participants with better AI scores	Reduction in time taken to complete task by participants the same participants after AI training
Unit test writing	35.29%	65.00%
Modernising code	64.71%	47.62%
Feature build	47.06%	65.00%
Code conversion	58.82%	72.22%

Our conclusion

Put simply, if you want your people to work faster, take the time to train them.

However, we can also deduce from our findings that the mastery of the AI tools is really short. So, instead of it taking a matter of weeks or months for people to perform tasks more efficiently and effectively, AI enablement can in fact leverage better results within a working day - if it's coupled with an investment in extra training and practice.

Lesson two:

Generative AI improves the quality of product engineering tasks.

The hypothesis:

AI can help our people complete key product engineering tasks to a higher quality

We're known for the quality and robustness of the products we build for our clients.

So working faster would mean very little if we weren't also delivering quality products.

More and more studies are finding that time pressure can improve productivity – but, under the wrong conditions, it also leads to decreased quality.

Which means we're constantly looking for that perfect balance between speed and value.

Previous experiments suggest that generative AI can improve product engineering quality in a number of ways, including:

- Suggesting and auto-completing code, reducing the chances of human error
- Picking out potential bugs or vulnerabilities in code rapidly and with greater accuracy
- Automating repetitive, lower-value tasks, giving users time to think more innovatively and iterate more rapidly

Results from previous studies were promising. In a 2024 McKinsey study, three out of four product managers found that using AI improved the quality of their deliverables.

Equally, we expected that speeding up product engineering tasks might also automatically improve quality, since developers who report fast code turnaround times feel 20% more innovative than developers who don't.

Faster and better:
Using generative AI helped our people deliver higher-quality work

Task	Average improvement in quality score
Unit test writing	51.00%
Modernising code	36.62%
Feature build	36.62%
Code conversion	41.82%

Once again, we’re seeing the potential power of generative AI. On average, independent evaluations found that, when participants used AI, the tasks were completed to a higher standard.

For engineers, the impact was particularly notable in some areas. For the unit testing task, 70% of engineers achieved a higher quality score using AI, compared to 23% who received a lower quality score.

Once again, “intuitive” generative AI users see greater results

Task	Average improvement in quality score	Average improvement in quality score seen by participants with better AI scores
Unit test writing	51.00%	70.59%
Modernising code	36.62%	64.71%
Feature build	36.62%	47.06%
Code conversion	41.82%	47.06%

It's relatively easy to understand that people who are comfortable using a tool will work more rapidly than those who aren't. They're likely to spend less time struggling with the tool itself, and more time working on the actual problem.

But the data clearly shows that understanding how to use AI effectively doesn't just help users complete tasks faster. It also helps users deliver higher-quality outputs.

Of course, speed and quality are inherently intertwined in product development. Higher-quality product development processes help users make fewer mistakes, which means fewer iterations are required – further multiplying the speed benefits we covered in the previous section.

...And generative AI training improves quality scores still further

Task	Average improvement in quality score	Average improvement in quality score after AI training
Unit test writing	51.00%	51.56%
Modernising code	36.62%	44.25%
Feature build	36.62%	44.25%
Code conversion	41.82%	40.74%

In this area, training had a more modest impact on quality, but we can still see a decided improvement in most areas.

Once again, there is one outlier: Code conversion quality actually dropped very slightly when participants were given further training. Further research will be needed to understand the exact cause of this, but as with code modernisation, human factors may still have a part to play in bringing this outlier about.

Our conclusion

Generative AI improves the quality of key product engineering tasks – and training in generative AI tools can improve quality further still.

The research shapes our methodologies

People are at the heart of AI

This research was never meant to be purely scientific.

We used ourselves as “guinea pigs” so that we could build a methodology based on facts, not hype.

To make sure our AI implementation has a real impact.

As a result, we can now pave the way for our clients’ own AI journeys. Not just by increasing the number of people using AI, but maximising the value that AI brings to an organisation.

Our research has taught us three core lessons. Lessons that we’re already putting into action with some of our clients – and that are helping them use AI to improve their effectiveness, create new efficiencies and make their operations smoother.



The lesson:

If you're not training your people to use generative AI, you're wasting your investment...

Or at least, you're not getting as much ROI as you could be.

Our research clearly shows that providing dedicated training – even a small amount of it – measurably improves task quality and the speed at which tasks are completed.

In other words, you can't give people generative AI tools and leave them to work it out for themselves. You need to make a conscious effort to help them develop new skills. After all, they'll be fundamentally changing how they work.

...So we're putting people and training at the heart of our generative AI strategy

For every one of our clients, the pathway to AI implementation starts with people. Preparing them to adopt AI by giving them the training and support they need to understand these new tools and use them to their full potential.

At CreateFuture that means:

Creating an "AI for all" training pathway

We've created a baseline understanding of AI across our entire company so that everyone has the skills they need to use basic, generalised generative AI tools.

Our training covers what generative AI is, our security policies for AI and how to use tools effectively and safely.

Our research shows that even a small level of this kind of general training can have a powerful impact, improving time savings and quality.

Specific role-based AI training

We're taking education even further with specific AI training that's designed to help key roles across the business improve their performance.

This involves some basic training, but it also involves using generative AI to target specific challenges on specific projects. Our people aren't just learning to use AI, they're learning how to problem-solve with AI – which means they're constantly pushing the boundaries of what generative AI can do.

The lesson:

Self-taught users can be powerful assets...

From the very beginning, we knew that most businesses have at least one employee that's already using AI. Even if it's in an unofficial capacity.

These self-taught AI frontier-crossers are using generalised tools like ChatGPT or Gemini to speed up their processes or improve their effectiveness.

That's a scary prospect for many business owners. And it's true that, if used incorrectly, generative AI can harm work quality – or even risk your business' reputation.

But our research shows that these “intuitive” generative AI users – those who naturally pick up AI tools with little training or guidance – can be a powerful asset.

A distinct group of participants in our research quickly adapted to using AI tools, even with no prior training. And these “intuitive” users consistently worked faster and delivered higher-quality work.

...So we're harnessing them as AI power users

Instead of forcing these users to follow strict training guidelines, we're turning them into sources of information and inspiration for the rest of the company.

These “intuitive” AI users are becoming AI champions and AI practice leads. They're running Lunch & Learns, speaking at external events and inputting into AI workshops. In other words, they've become an extra, experienced voice in the AI adoption process. A bottom-up supplement to our top-down training strategy.

The lesson:

There's still much more to learn...

Our research proved some of our hypotheses. But it also threw up new questions.

Why did users who received training save less time modernising code? And why did training seem to reduce code conversion quality?

Equally, some areas will need further exploration. How do different types of training affect quality and speed? How do we train people enough to save time and improve quality, without wasting time that could be used on other tasks?

The world of AI is still generally unknown. And, as new tools enter the market and new approaches are discovered, the horizon of that world is constantly shifting.

Which means our AI strategy can't be set in stone. It needs to be constantly evolving.

...So we're constantly testing and evolving

We're taking a "kaizen" approach to AI implementation – constantly learning, monitoring our impact and adjusting our approach. Making small adjustments and improvements every day that add up to major advances over time.

We'll continue to be our own guinea pigs. We're already using the lessons we've learnt from our own AI implementation to shape strategies for our clients. And each of those projects teaches us something else about AI adoption, which we'll feed into the next project.

It's a constant process – but a constant climb towards perfection too.

Creating tomorrow, together, today: Powered by people, accelerated by AI

The insights we gained from this study now shape our entire approach to AI at CreateFuture.

We don't want to just build AI-enabled products. We're much more excited about using AI – or helping our clients use AI – to unlock new efficiencies, optimise processes and do better work.

That's why we put people at the heart of everything we do.

We're becoming a digital consultancy that's powered by people and accelerated by AI.

And, today, we're helping our clients unleash their potential by leveraging our expertise in software engineering, innovation, strategy and design to create digital products that solve their challenges. Using generative AI as a tool to reach higher, go further and create real change.

Want to find out more about how we can help you embrace generative AI? Get in touch with our team [here](#).



CreateFuture

CREATEFUTURE.COM