Principles in Action
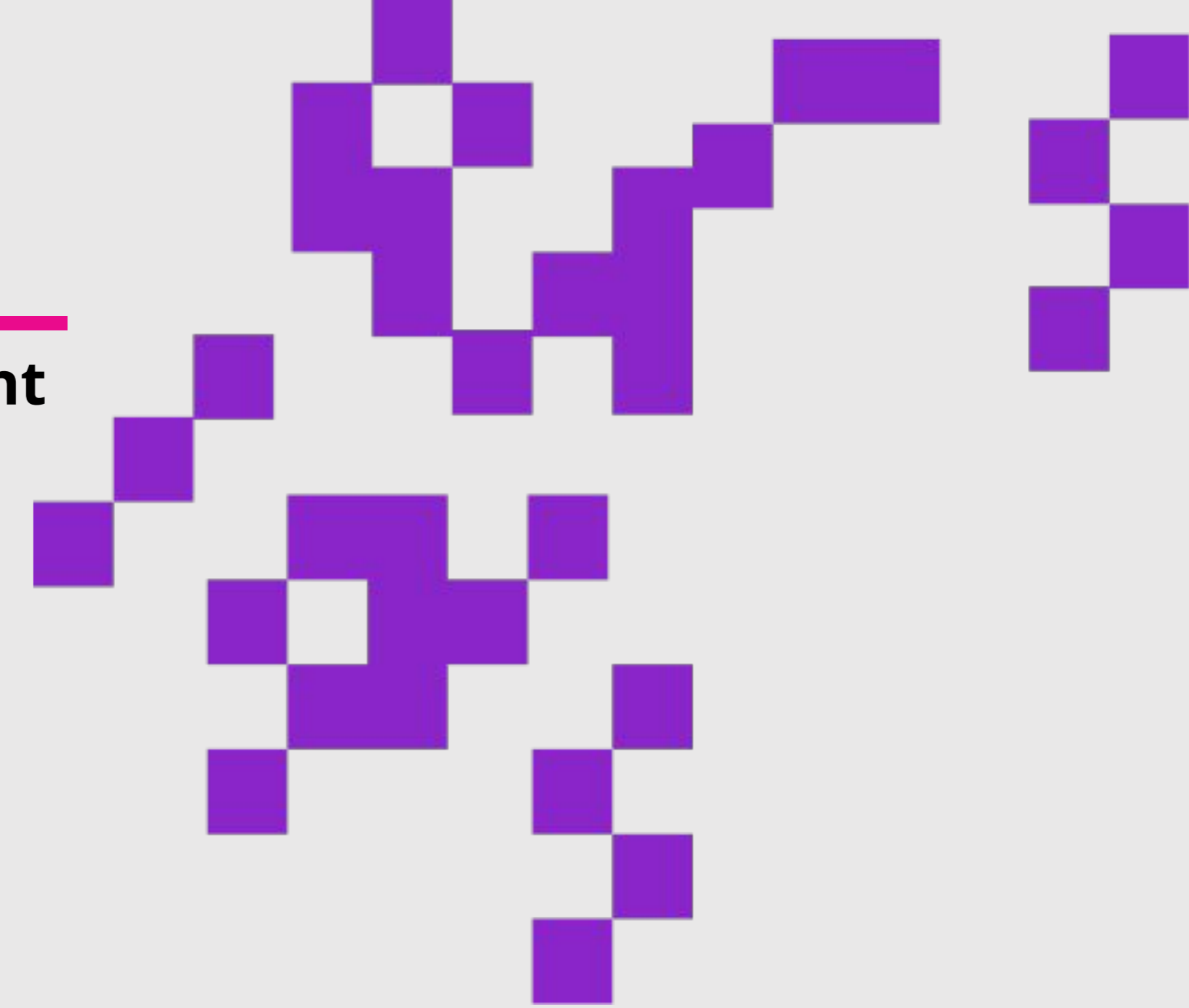
A Playbook for Responsible AI
Product Development

Welcome

The Principles in Action Playbook contains guidance, examples, and tips to help those building products that leverage AI to do so responsibly. It was created to help professionals on the front lines - entrepreneurs, product managers, designers, tech leads, and domain experts - navigate the complexities of AI product development.

The Playbook is a resource that you can use to ensure that the people, infrastructure, time, money, and energy that you invest into AI development can make users happy and make this world better. We hope you'll use the playbook not only to design AI in a safe, trustworthy, and ethical way, but as a means for building a culture around human-centred AI within your teams and organization.

The playbook sits in an ecosystem of resources that already exist. We're not the first piece of guidance around AI that other research institutes, governments, or AI industry leaders have put out. On one side, we have Vector Institute's AI Trust and Safety Principles - these are what guide every major decision we make. On the other side, we have tools like UnBIAS and CyclOps that solve some of the biggest challenges for AI developers, such as fairness and evaluation. The playbook is a connection between the two; it's a way to put the AI principles in action.

Ideation — **Development** — **Testing** — **Deployment** — **Post-Deployment**

The playbook has five stages that cover the product development process - from coming up with ideas for your product, to building your AI model, to testing your product before deploying it, all the way to what you should do after you launch.

We encourage you to start by skimming through each stage and reflecting on the questions, especially if you're already hip-deep in development, to assess where you're at and where you still need to go.

Ideation

Should you use AI?

1. Are you building something good for the world?

2. Do you need to use AI?

3. How realistic is your solution?

4. What are your success metrics?

5. How can you mitigate risks?

1. Are you building something good for the world?

We're excited that you're looking to build something that leverages AI. As you go through the product development process, focus on being people-first over technology-first.

For example, instead of approaching your brainstorming with "How can we use AI to ___?"
➔ Move to "How might we solve [human need]?"
➔ And then "Can AI solve [this need] in a unique and helpful way?"

'**How might we**' statements force you to start with a human need and steer you away from suggesting a solution, so that you can be open to generating further possibilities.
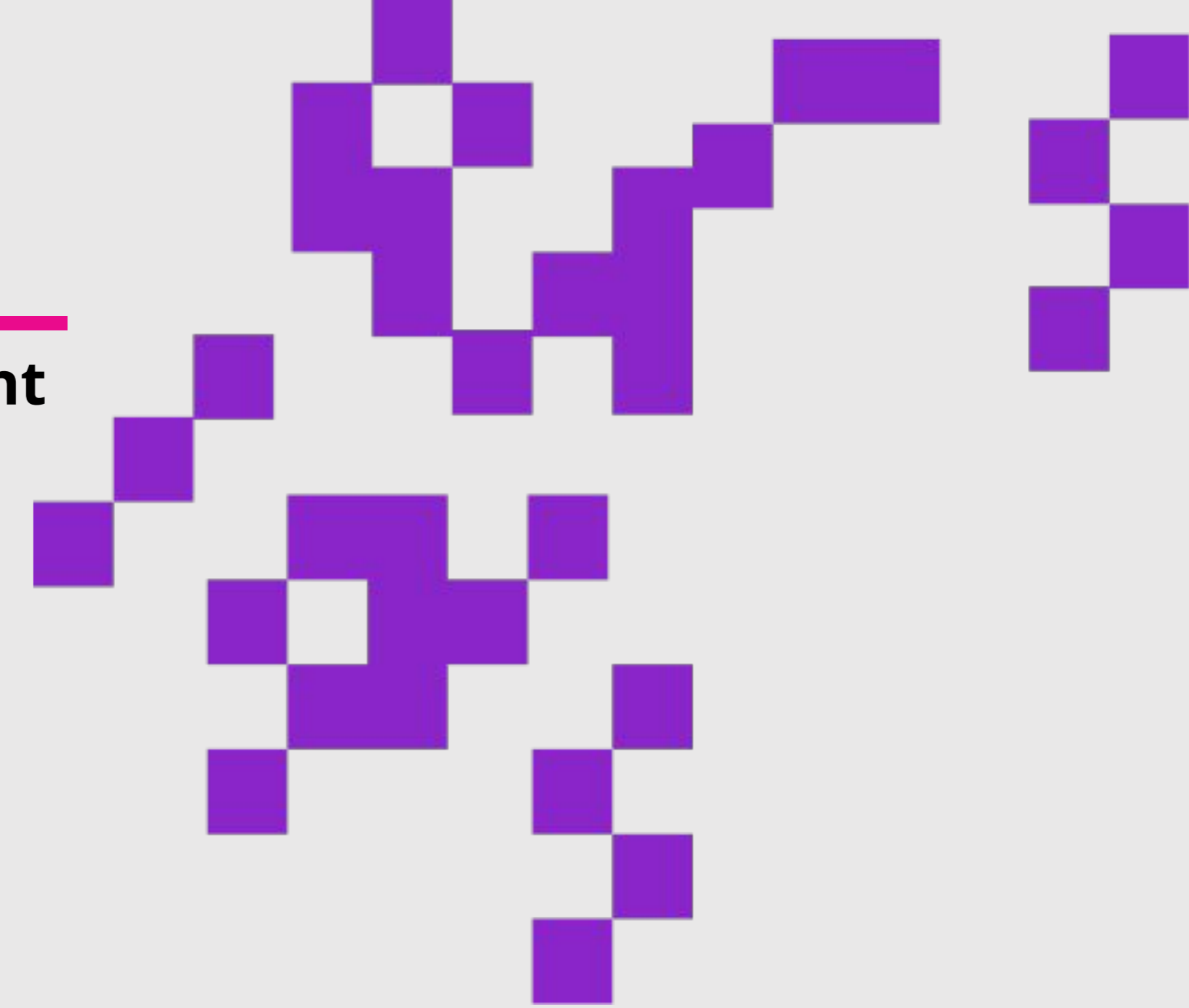
How do you decide what problem you should tackle?

1. **Identify a need humans have.**
   a. Engage with a diverse group of people and scenarios.
2. **Collect evidence that this is a problem.**
   a. Listen to people.
   b. Look at data.
   c. Watch behaviours.

2. Do you need to use AI?

Don't use AI just because you can. Think hard about whether you really need AI in your product to solve your problem:

1. Think through the journey of the need, task, or user experience you want to improve.
2. Using the chart below as a starting point, consider whether adding AI **will improve your product experience, do nothing for your product, or maybe even degrade it.**
3. Interview potential users to understand their concerns and consequences of using AI. Round tables and focus groups work well for this.

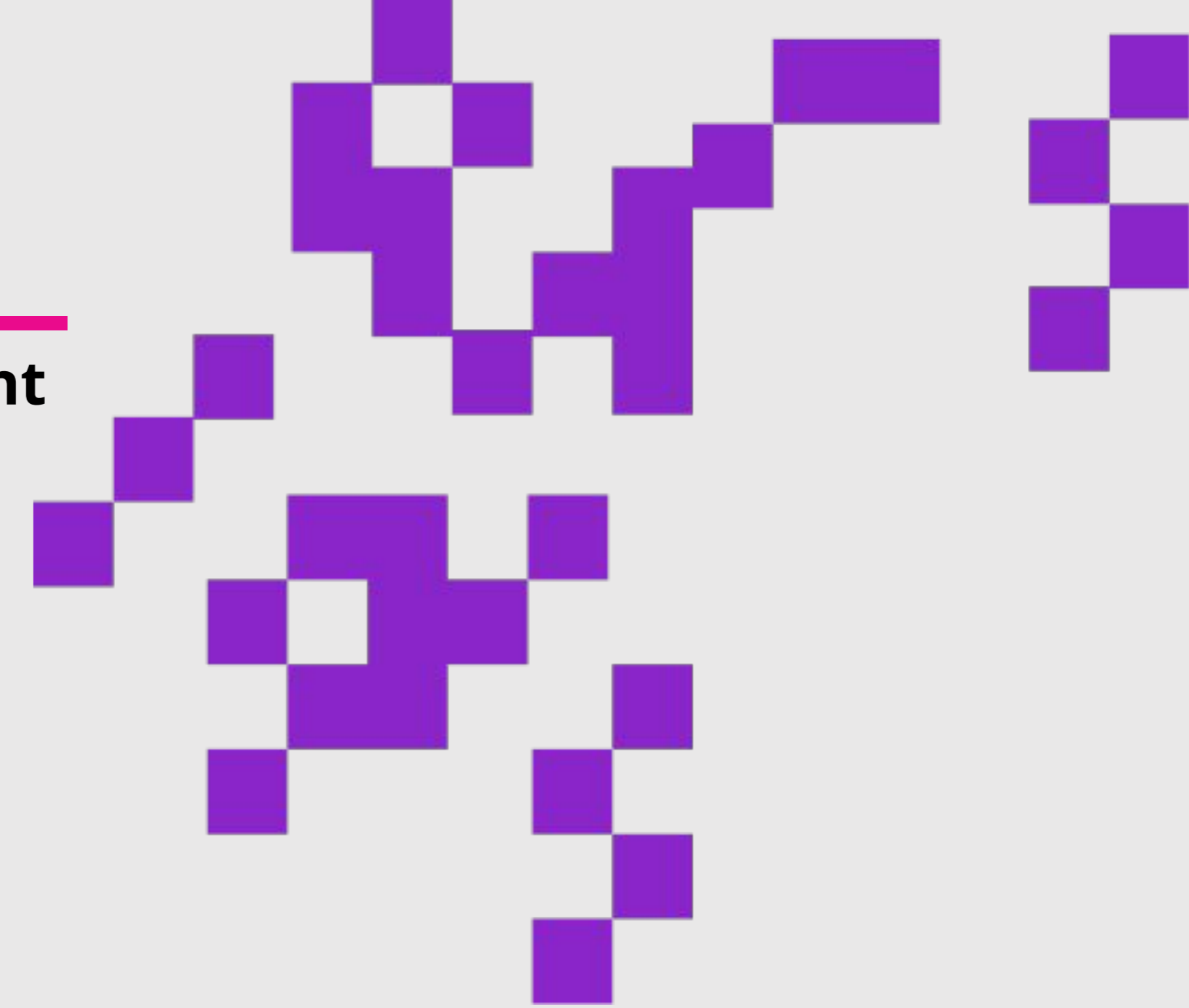| **AI is good at**	**Avoid AI when**
• Automating tasks that people don't know how to do, or find boring, repetitive, or dangerous	• People tell you they want to do a task without help
• Augmenting tasks that people enjoy doing	• People want creative control to see their vision through
• Recommending different content to different users	• Being predictable is essential, always
• Predicting and forecasting trends and events	• The cost of errors is greater than the benefits of a small increase in success rate
• Personalising a user experience	• You and your customers need to understand exactly why something happened
• Understanding human language	• Shipping fast is priority
• Recognizing patterns in images, text, or numbers	• Solving novel situations where there is limited data available for training
• Detecting anomalies	• Making ethical decisions
• Generating customized text, images, or music	• Giving users manual control is a better user experience
	• A rule-based solution will get the job done
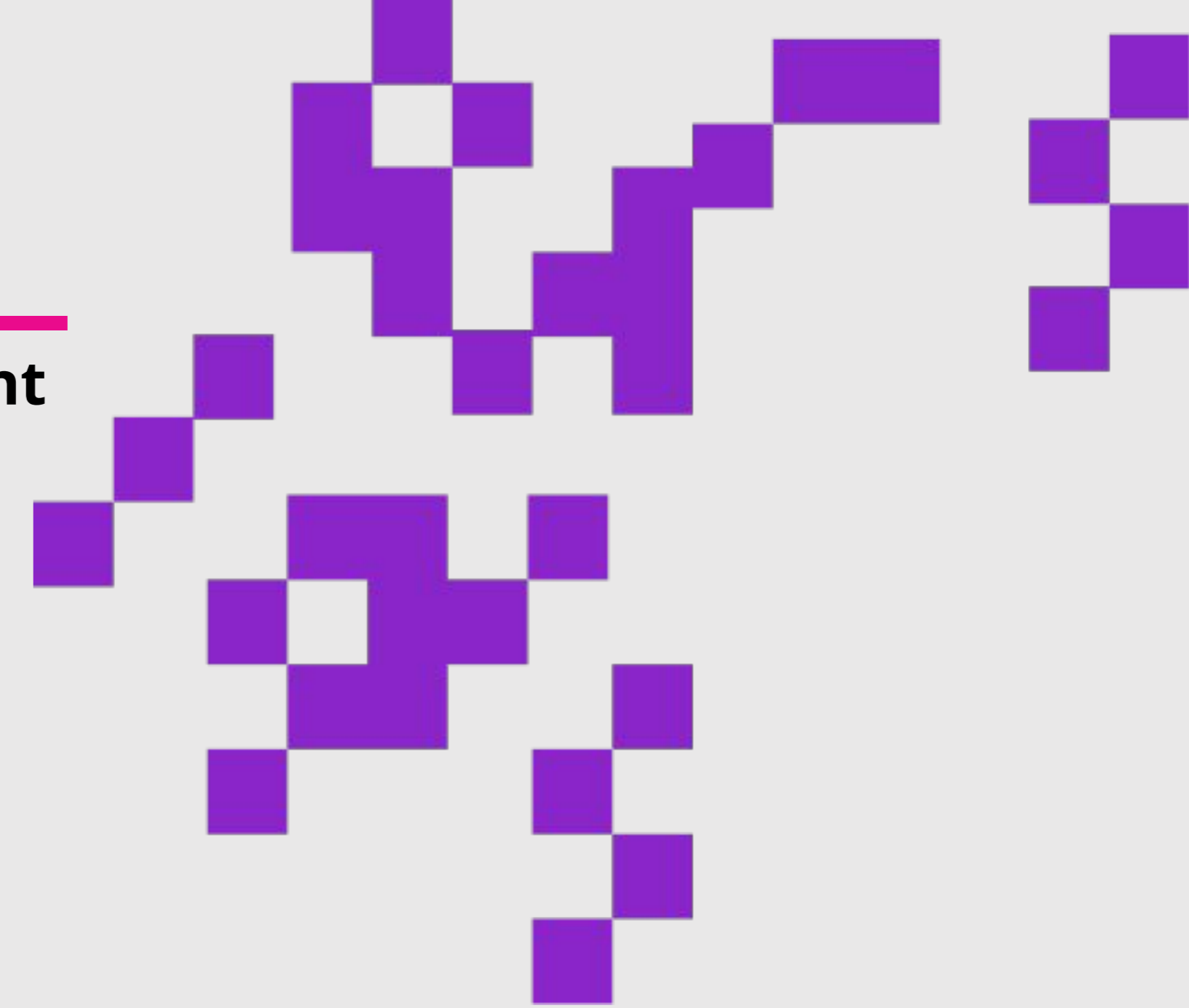
If leveraging AI makes sense for your product, then consider this:

- Know your user, know who is not your user. Know how they'll both be impacted.
- Know who might be indirectly impacted, including individuals, communities, organizations, society, and the planet.
- Consult marginalized populations and public safety groups.
- Choose the appropriate level of automation for the task; consider what is at stake for the user, and how comfortable they would be handing over the task to a system.
- User has complete control over how to proceed with the prediction.
- User is presented with suggestions on how to proceed.
- The system chooses how to proceed on behalf of the user.

3. How realistic is your solution?

Evaluate your capacity before you start building to decide whether your team is able to implement your idea.

- Diverse team members who have the technical expertise to build, evaluate, and deploy AI products.
- Subject matter experts who can commit to being a part of the entire product development process and share insights.
- Good quality data because your AI is only as good as the data that feeds it.
- Time to devote to a lot of experimental work with potentially no satisfactory results.
- An understanding of applicable local and international laws and regulations, including those related to human rights, data protection and privacy, security, copyright, and intellectual property.
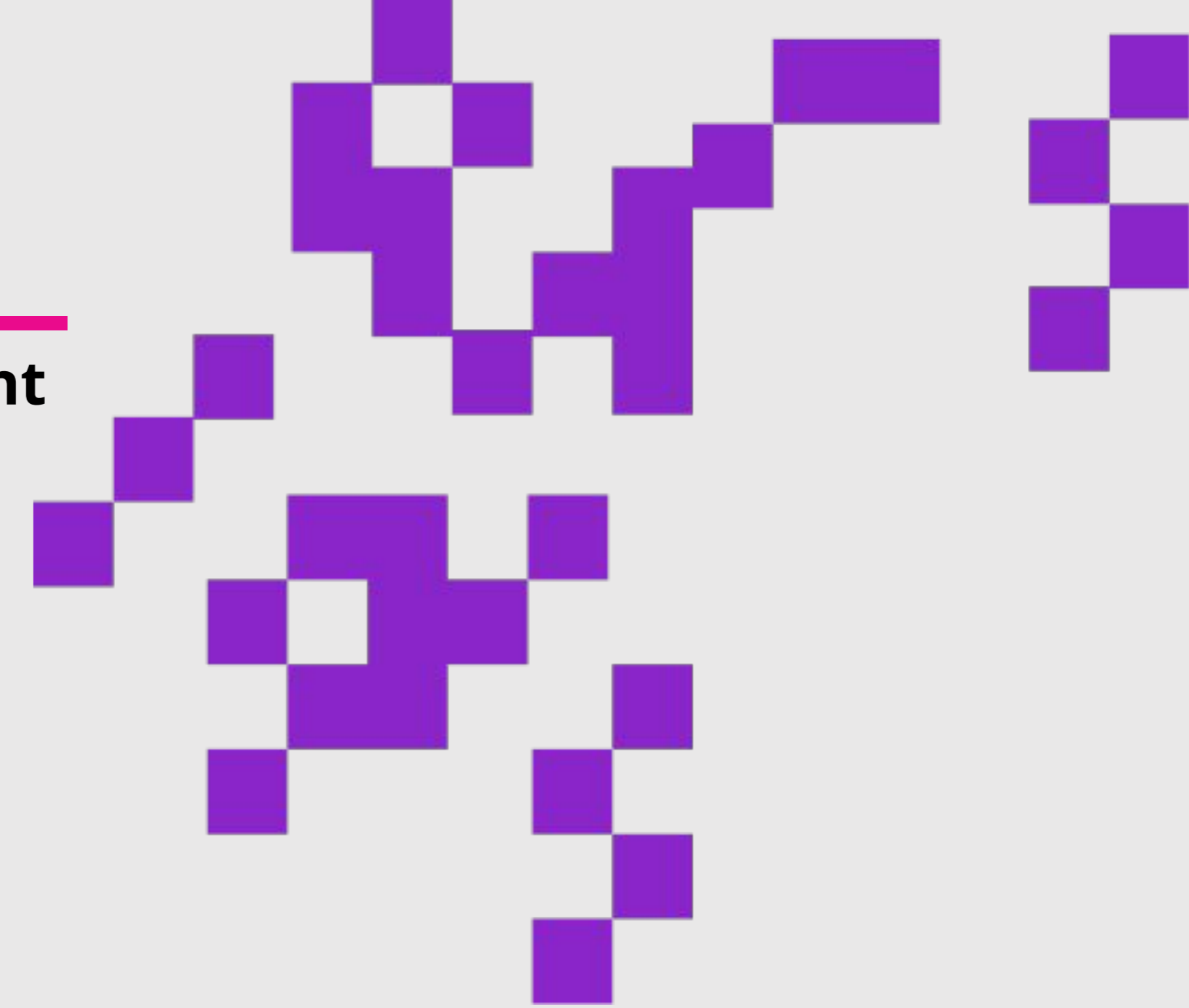
4. What are your success metrics?

How will you know when your AI system is good enough for people to use?

Start by identifying the action or behaviour you are trying to optimize and the possible outcomes. If your AI product will make predictions, you can expect successes: true positives and true negatives, and errors: false positives and false negatives.

Think through the consequences of false positive and false negative predictions and weigh the cost of these errors.

Next, consider how your **model metrics** translate to your **product metrics**. When choosing high-level product metrics, such as engagement, speed, or cost savings, consider the following:

1. Choose several metrics to track, including proxy metrics and counter metrics, rather than a single metric to evaluate trade-offs between different types of outcomes, including benefits and harms. This will give you a broader picture of how your system is performing.
2. Have baseline measurements to compare against.
3. Be able to slice your metrics across user subgroups. You need to know if your feature is benefiting all user types or negatively impacting some people.
4. Meaningful feedback, especially at the initial launch, can be collected through user surveys.
5. Assign someone responsible to drive and report the metrics.

5. How can you mitigate risks?

AI is not perfect; it's probabilistic, so you should expect your product to give users incorrect or unforeseen output at some point, and those consequences can have their own consequences, also known as second-order effects.

Plan to design your user experience around these error possibilities.

1. Ask your users what they expect in different scenarios and what they feel is a success for the action or behaviour you're trying to help them achieve.
2. Explain the AI system's output to your users.
3. Design a "human-in-the-loop" experience. People should have ways to preview, review, edit, undo, dismiss, ignore, or manually take over control.
4. Give users a way to provide product feedback and report issues.
5. Monitor your model performance metrics on a regular basis and set benchmarks for certain actions that need to be taken when a metric drops below a certain threshold.
6. Take an ethics-focused approach to data protection and privacy, terms of use, and customer education.
7. Discuss how you will mitigate bias, avoid discrimination, and ensure fairness.
8. The default reaction to poor AI system output doesn't always have to be to fix your AI model to get better results; you can make design changes to the user experience, too.
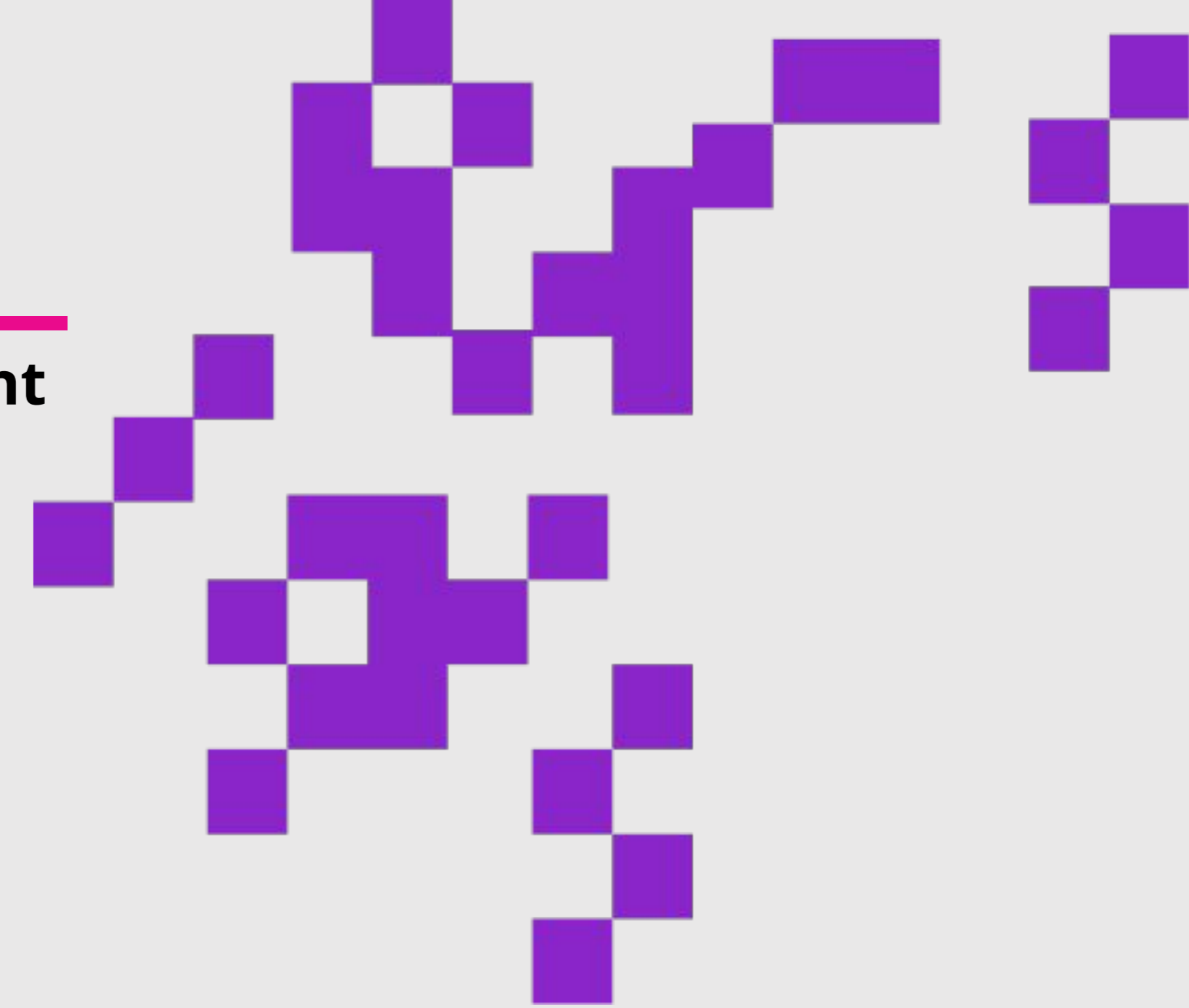
Development

How do you build something with AI?

1. Do you need to build your own AI?

2. Where should you get your data from and how do you evaluate it?

3. How do you build a responsible model?

1. Do you need to build your own AI?

Building AI solutions differs from traditional software development where there are often defined product milestones, requirements, and estimates.

Whether or not you decide to build your own AI depends on various factors such as your specific goals, resources, expertise, and the availability of suitable AI solutions in the market.

Reasons to build your own model	Reasons not to build your own model
• Automating tasks that people don't know how to do, or find boring, repetitive, or dangerous	• Investing in intensive research and exploration is not a priority
• Augmenting tasks that people enjoy doing	• Sourcing sufficient high quality data for model training and testing will be challenging
• Recommending different content to different users	• Your team does not have the expertise and resources in machine learning, data science, and software engineering
• Predicting and forecasting trends and events	• You do not have the budget to support infrastructure costs and ongoing maintenance
• Personalising a user experience	
• Understanding human language	
• Recognizing patterns in images, text, or numbers	
• Detecting anomalies	
• Generating customized text, images, or music	
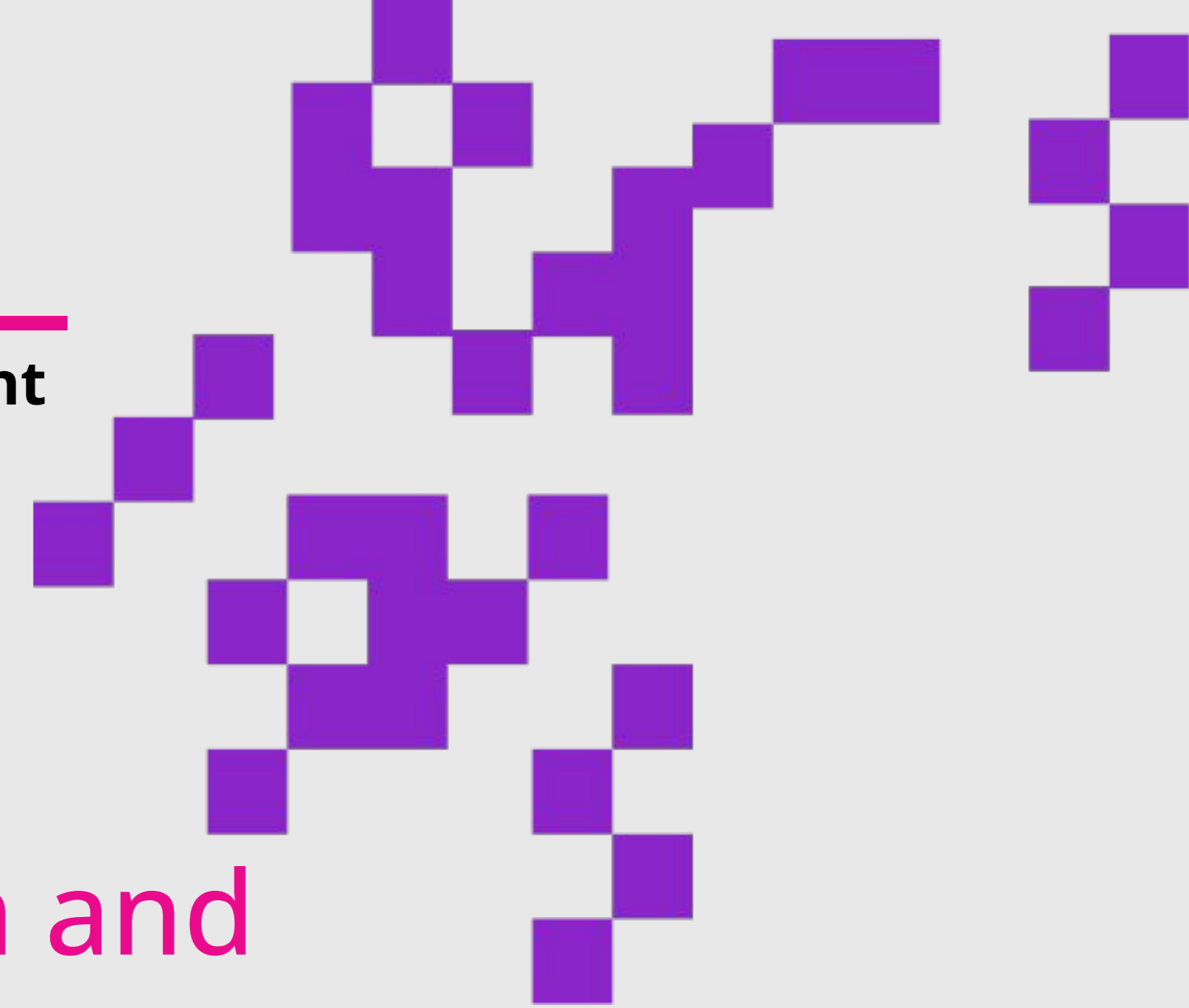
# 2. Where should you get your data from and how do you evaluate it?

If your team has decided to build its own AI model or fine-tune an off-the-shelf solution you will need data for training and testing. Your AI model, and thus your product, will only be as good as the data and labels that feed it, so think through your data needs carefully.

Remember, it is your responsibility to minimize unfair bias in your dataset. You are not absolved of your human rights obligations just because you don't have access to good data.

Consulting people with subject matter expertise will greatly help you in this process. **Domain experts** don't need to be data experts, they just need to be willing to share insights and highlight implications about your data's subject matter.

Here are some factors to consider when sourcing your dataset:

- Do you have a data card that documents information about your dataset?
- Was the data collected responsibly?
  - Do you require a licence to use the dataset?
  - Have you attributed or cited the publisher of the dataset as a way to acknowledge their work and give users the ability to re-use or fact check the data?
  - Are you legally allowed to use and store the data in the geographic regions you plan to release your product?
- What preprocessing has the data gone through?
- How closely is the dataset representative of your users and your use case?
  - Does the data reflect the real world? Consider user demographics, recency, time of year, trends, global events, image quality, and mistakes in text.
  - Is the data noisy? That is not necessarily a negative; allowing for imperfect data can more closely match the data you will get from your user base.
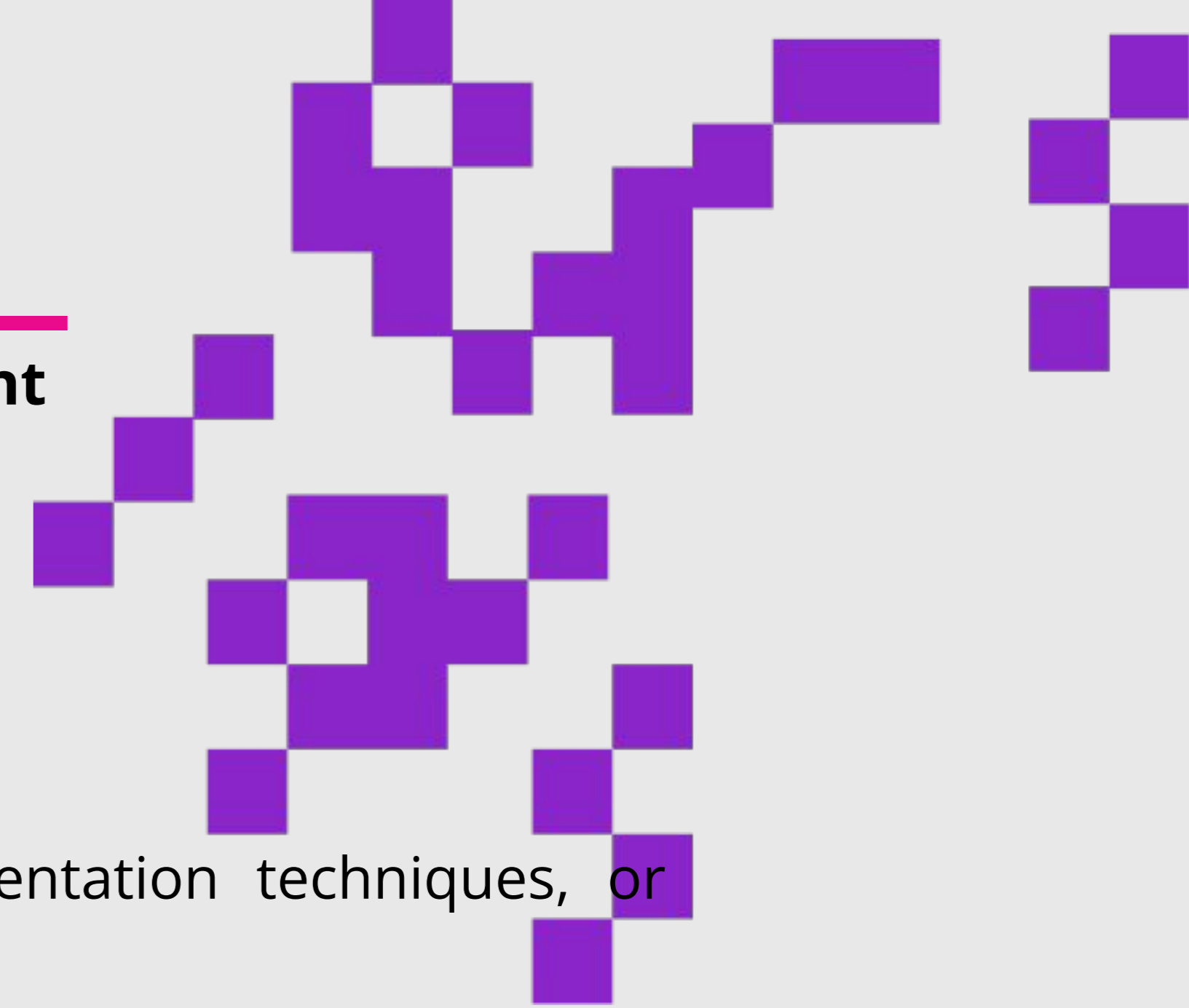
- ○ Do you need to modify the dataset with additional data or augmentation techniques, or combine multiple datasets?
- ○ If you can't find representative data, are you comfortable limiting your product release to only demographics of users reflected in your dataset, provided you're not discriminating against a protected group?
- Does the data source have known biases?
  - ○ Is your data collection method biased?
  - ○ How have you minimized unfair bias in your dataset?
- What is the quality of your data labelling? Hiring data labellers isn't cheap, and neither is quality labelling, especially when it's done ethically.
  - ○ Have labellers been trained on identifying and mitigating bias and discrimination to accurately label data?
  - ○ Have you given labellers sufficient guidelines to label and agree on labels?
  - ○ Have you set up a labelling process that compensates labellers fairly, ensures safe working conditions, and respects workers' ethical boundaries on sensitive data?
  - ○ Should you train domain experts to create gold standard labels?
  - ○ Are there mistakes? Scrub the data for missing values, duplicates, inconsistent formatting, or incorrect labels.
- Is the dataset adhering to privacy and security standards?
  - ○ Have you redacted all personal identifiable information?
  - ○ Have you aggregated data to maintain anonymity?
  - ○ Do the right people have permission to access the data securely?
  - ○ Is data encrypted and stored securely?
  - ○ Have you considered privacy methodologies - including differential privacy, federated learning, homomorphic encryption, or synthetic data generation – to address privacy risks and mitigate them?
- Are you able to maintain your dataset going forward?
  - ○ Do you have manual or automated data inspection and quality assessment mechanisms to ensure the quality of data?
  - ○ How will you know when data is outdated?
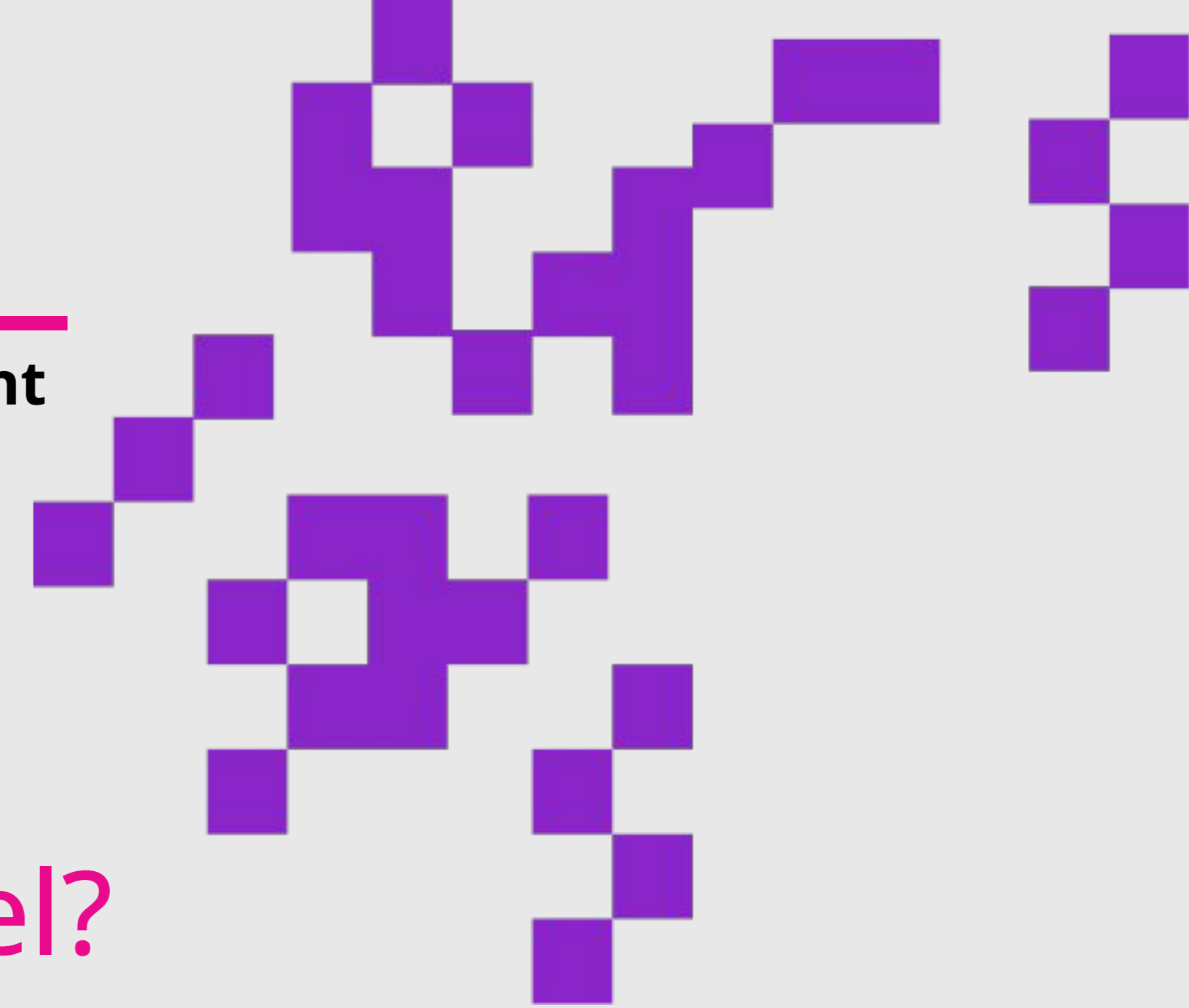
3. How do you build a responsible model?

Now that you have good quality data that reflects your users and use case, you can start thinking about how to develop the model that will output predictions or content that will help address your users' needs.

Here are some tips:

- You don't need to use the latest machine learning models. Most users aren't concerned about which state of the art model you're using, only that they are getting the information they need to get their task done.
- In addition to biases in data, AI systems can behave unfairly because of societal biases that are either explicitly or implicitly reflected in the decisions made by your team during model development.
- Prioritize explainability and interpretability of your model, especially when decisions impact human well-being, such as healthcare, employment, justice, or finance. People affected by these decisions deserve to understand the reasoning behind them.
  - If possible, add features to your model one-by-one to determine the impact of different features and how they affect model performance.
- Train the model to be robust against adversarial attacks.
- Ensure that your model is trained on a secure network to protect data and access.
- Training, maintaining, and running inference on machine learning models consume significant energy and environmental resources. The carbon emissions of AI systems can vary depending on factors such as the specific tasks they perform, the type of data they process, and the hardware used. To reduce your footprint, consider creating smaller versions of trained models, improving the efficiency of your operations, or participating in a carbon offset program.
- The idea of achieving an optimally responsible model is a fallacy; developing models is a constant balance of making trade-offs for your unique use case.
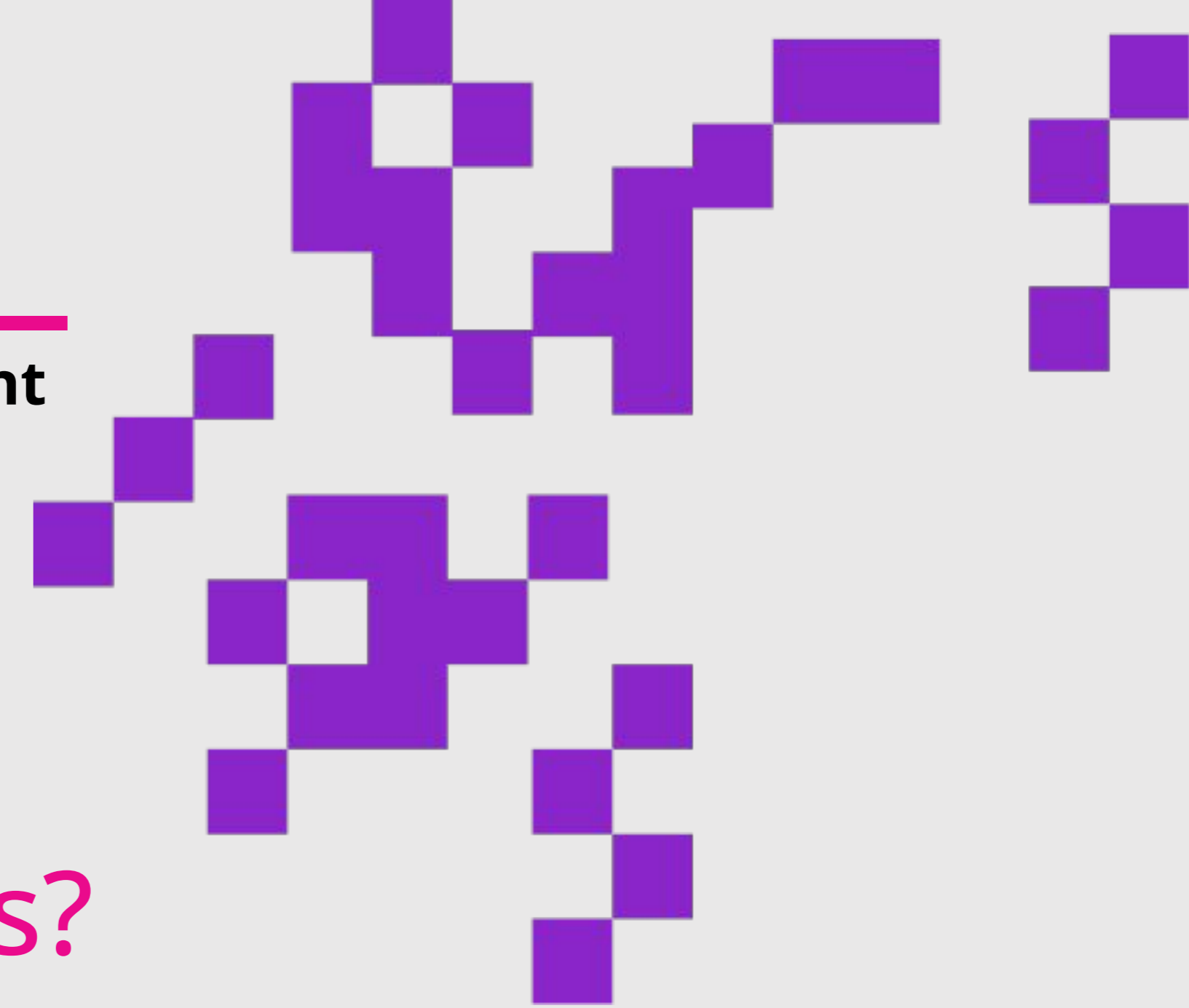
Testing

How do you know if your AI works?

1. How do you ensure model robustness?

2. Is there human oversight?

3. What metrics matter to you?

1. How do you ensure model robustness?

Test, test, test!

This is true in traditional software engineering, and is especially true for model development.

We can say with confidence that your AI system will give wrong and unexpected outputs at some point. To limit bad predictions users might encounter, **make a plan for testing** early on in your product life cycle.

Here are some tips:

- Anticipate the types of errors users might encounter and evaluate their consequences. What is the impact of a false positive and a false negative prediction for your use case?
- Continuously evaluate your system. Every time you push a feature change, whether of your model or even a design change to your UX or UI, A/B test the current system against the new system to verify the impacts. Testing isn't a one and done event.
- Ensure your AI system is regularly tested by a dedicated security team.
  - Consider obtaining a privacy audit to stay in compliance with relevant laws and regulatory requirements.
  - Conduct red-teaming to try and manipulate, misuse, or confuse your AI system so that you can uncover vulnerabilities to malicious actors or unintended behaviours. Focus on ethical concerns, safety, and fairness to ensure that potential harm to users or groups is mitigated.
- Revisit your data. Poor model output can often be traced back to poor data that you may want to clean again.
- Expect to set aside time to tune your model. Maybe even slightly overestimate the effort and time to do this task. Model development is an iterative process - make sure your product leaders and stakeholders understand that. Quick, positive outcomes achieved by implementing the happy path is an effective way to gain the favour of both your product manager and user.

- In use cases where the consequence of errors is high, such as in health care, perform a silent trial of your AI system, which evaluates your model on prospective data (e.g. patients), while the end users (e.g. clinicians) are blinded to predictions such that they do not influence decision making.
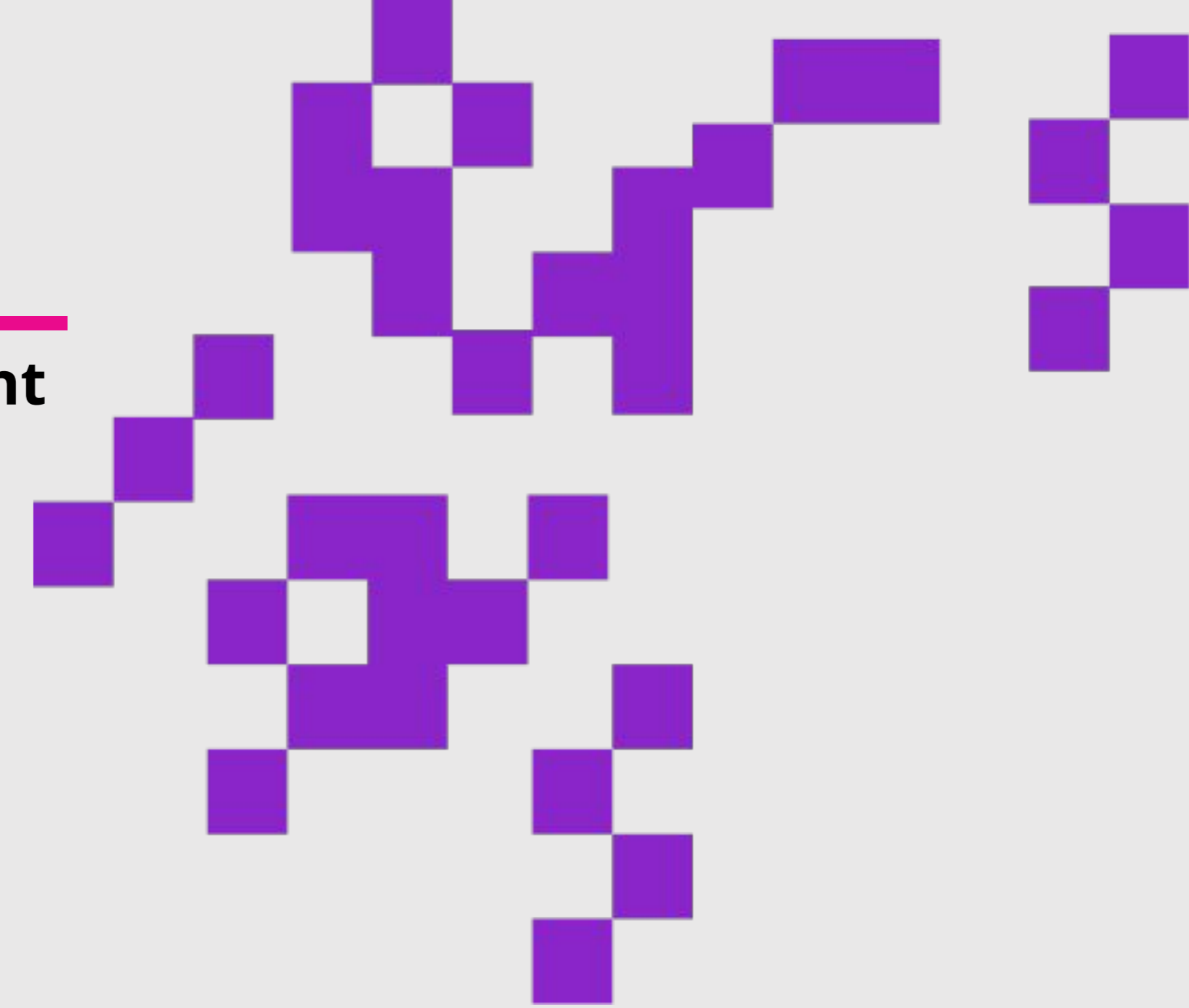- Do your unit tests and integration tests.

2. Is there human oversight?

Focus on initially getting qualitative feedback from a variety of users instead of obsessively tracking metrics. Your users will quickly tell you if it's not working as expected.

Perform **retrospective testing** on your model. Have multiple individuals with the appropriate expertise examine cases where there is a mismatch between the model output and the ground truth. Measure the agreement between humans and the model, and between humans.

Next, you can test using a **prototype** within your team, then within your company. Avoid relying on the development team for testing as they may only test particular aspects of the system or look to confirm something they already know. The test cases they write are typically limited to how they think the system will be used. By opening up testing to a variety of people in your organization, you can observe the variety of alternative ways the system is used by real people.

Eventually, bring in **actual people** to observe how they interact with your product. Test with a small group of representative users in a staging environment and give users the option to use or ignore your model's predictions. Pre- and post-surveys work well to understand how users felt about a product change. Quick, optional pop-ups presented at the time of output are ideal.

3. What metrics matter to you?

There are many metrics you can use to evaluate your AI model. But at the end of the day, what you really care about is assessing whether you've addressed your target user's needs in a responsible way. Therefore, the performance of your model should be measured against product success metrics and bias and fairness metrics. Choose metrics that are simple to measure.

For product metrics:

First, measure a user behaviour that is directly observed and attributable to an action of the system:

- Was this ranked item selected?
- Was this ranked item rated?
- Was this shown item marked as not useful/ offensive?

During A/B testing and launch decisions, measure indirect effects:

- Did the user visit the next day?
- How long did the user visit the app?
- What were the daily active users?

Finally, don't expect your new AI to tell you:

- Is the user happy using the product?
- Is the product improving the user's overall well being?
- How will this affect the company's overall health?

These are all important, but also incredibly hard to measure. Instead, use indirect indicators: if the user is happy, they will stay on the site longer. If the user is satisfied, they will visit again tomorrow.

Remember, there is no one metric that will tell you that your product is a good one and a responsible one. Your team should care about engagement, daily active users, retention, and revenue during A/B testing. But while A/B tests help us optimize specific elements, they're not the end goal. Your true focus is on creating a product that delights users, attracts more customers, fosters strong partnerships, and drives sustainable growth in a safe, trustworthy, and ethical way.

As such, always assess your product metrics with **fairness and bias** in mind:

- Expect to manage differing opinions across stakeholders on the value of fairness.
- Conduct a fairness analysis to evaluate how various subgroups of users are treated by the AI system.
  - Compare product metrics between different user demographics.
  - Get a better picture of how your model is performing across subpopulations by measuring AI fairness metrics such as equalized odds, balanced accuracy, and predictive parity.
- The goal is not to fully "debias" a system, but to detect and mitigate fairness-related harms, such as:
  - Unfairly allocating opportunities, resources, or information.
  - Failing to provide the same quality of service to some people as you do to others.
  - Reinforcing existing societal stereotypes.
  - Criticising people by being actively derogatory or offensive.
  - Over- or under-representing groups of people, or even treating them as if they don't exist.

One of the challenging parts of model development may be using your own judgement to assess trade-offs while optimising for certain metrics, e.g. fairness vs accuracy or false positives vs false negatives.
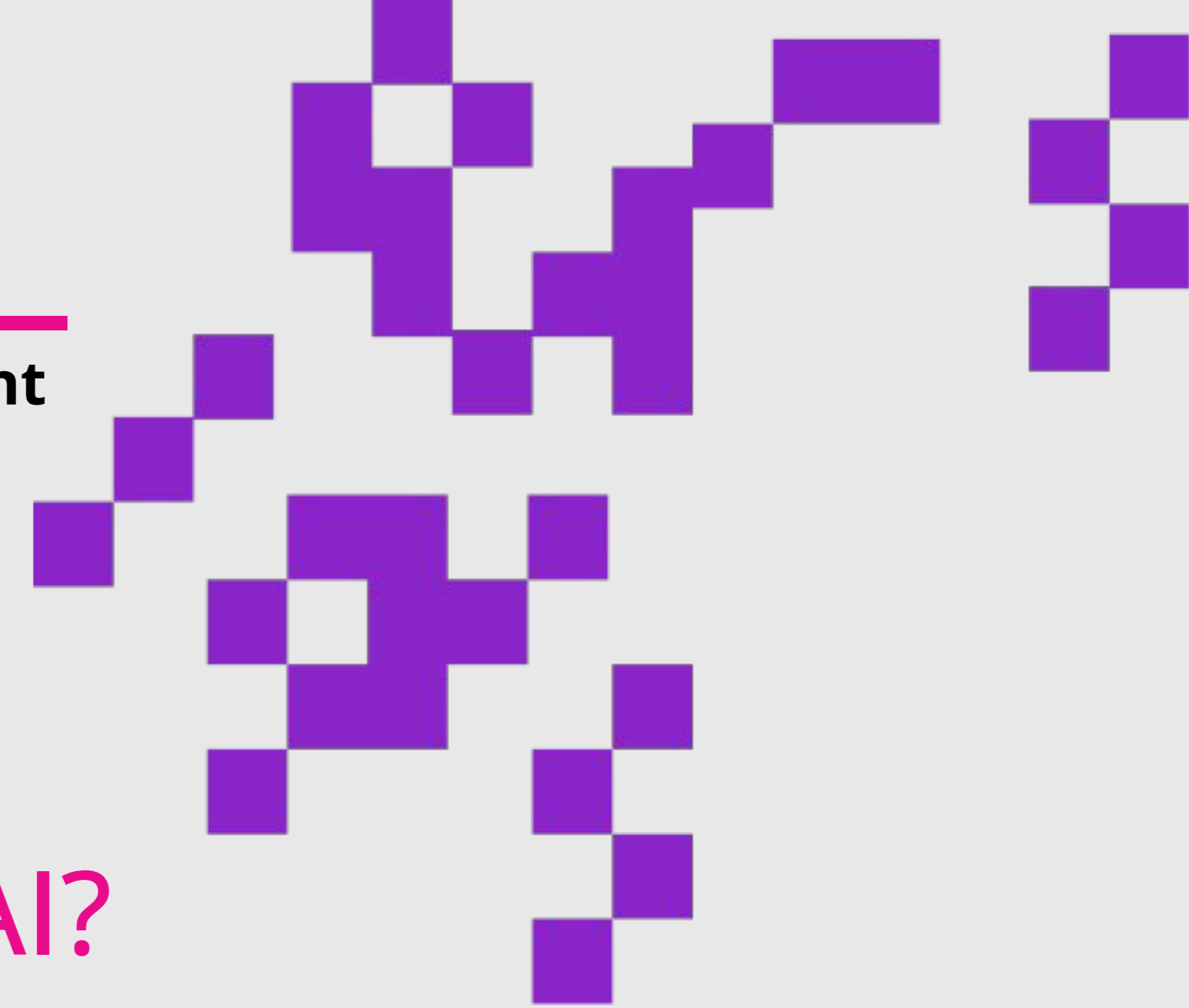
Deployment

How do you safely launch your AI?

1. How will you prepare your users for AI?

2. How transparent is your AI?

3. Do you give users options for personal control?

1. How will you prepare your users for AI?

Work on building **trust** from day one.

- Trust starts long before users use your product; it starts with the marketing material they consume.
- Notify users that your product uses AI.
- Be cognizant of the language you use to describe your AI-powered product. Messaging such as "magical" and "human-like" can leave users with the wrong impression of what your system is capable of.
- Build trust through others. If appropriate, leverage the users' networks as social proof.

Phase rollout.

- Start with a small group of diverse beta users, ideally ones who have opted to try your new product or feature, before launching to your entire user base. When you're confident in your system's output, expand to additional users in batches.
- Only launch if you've reached the model performance and product success metrics you've set for yourself or agreed upon with potential customers. Ensure the system is well-tested and has no major bugs before rollout to continue building user trust.

Prepare to **onboard** users to the new product or feature.

- Provide guidance on best practices, show the benefits of your product through demos or examples, give users a playground for experimenting with no consequences, add clear, brief tooltips or placeholder text to educate in-the-moment, and make help documentation accessible.
- Anticipate that some users will not complete your onboarding process and might get frustrated later on. Try to include an interactive practice scenario during their first session so that they can quickly experience the value-add.
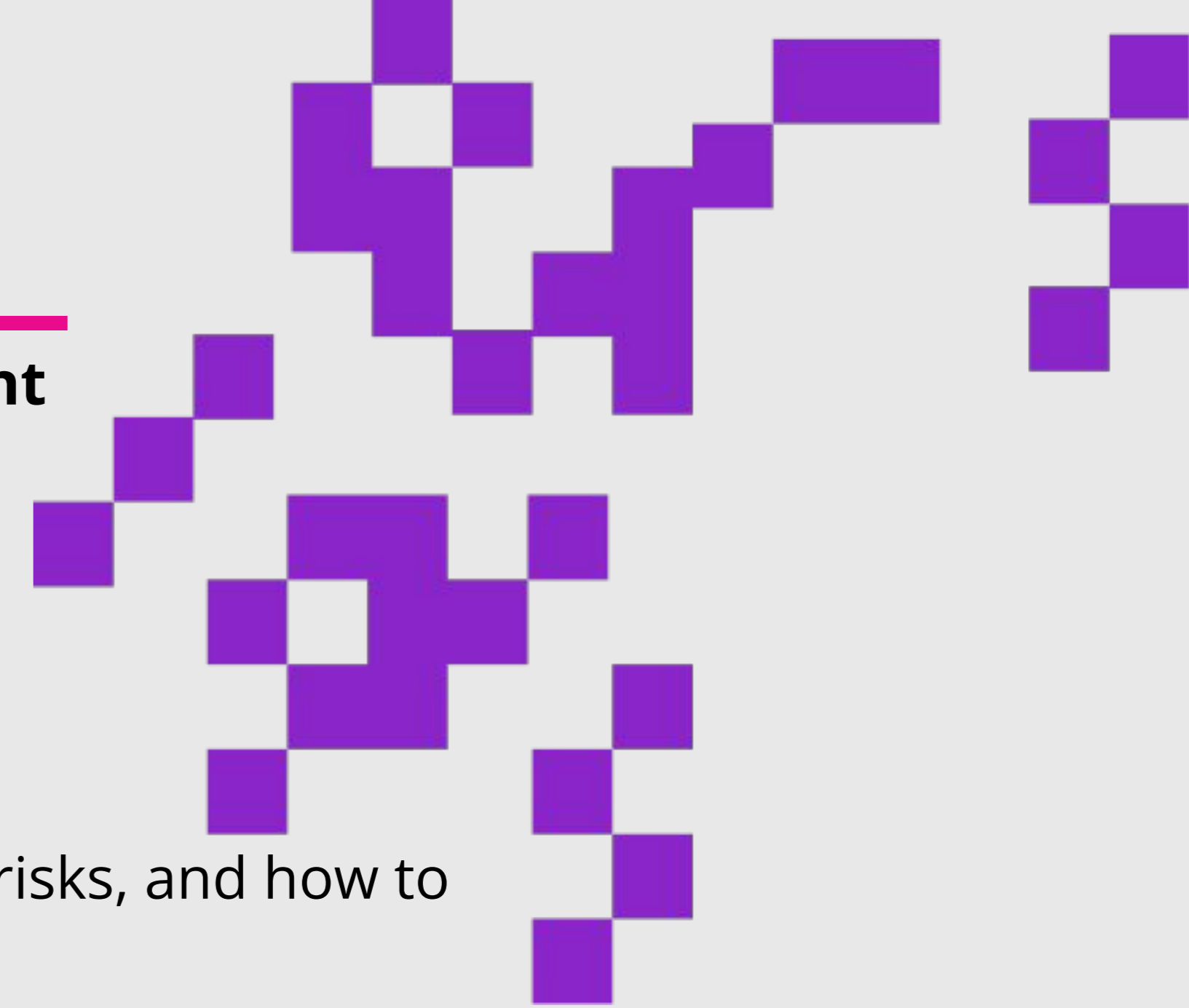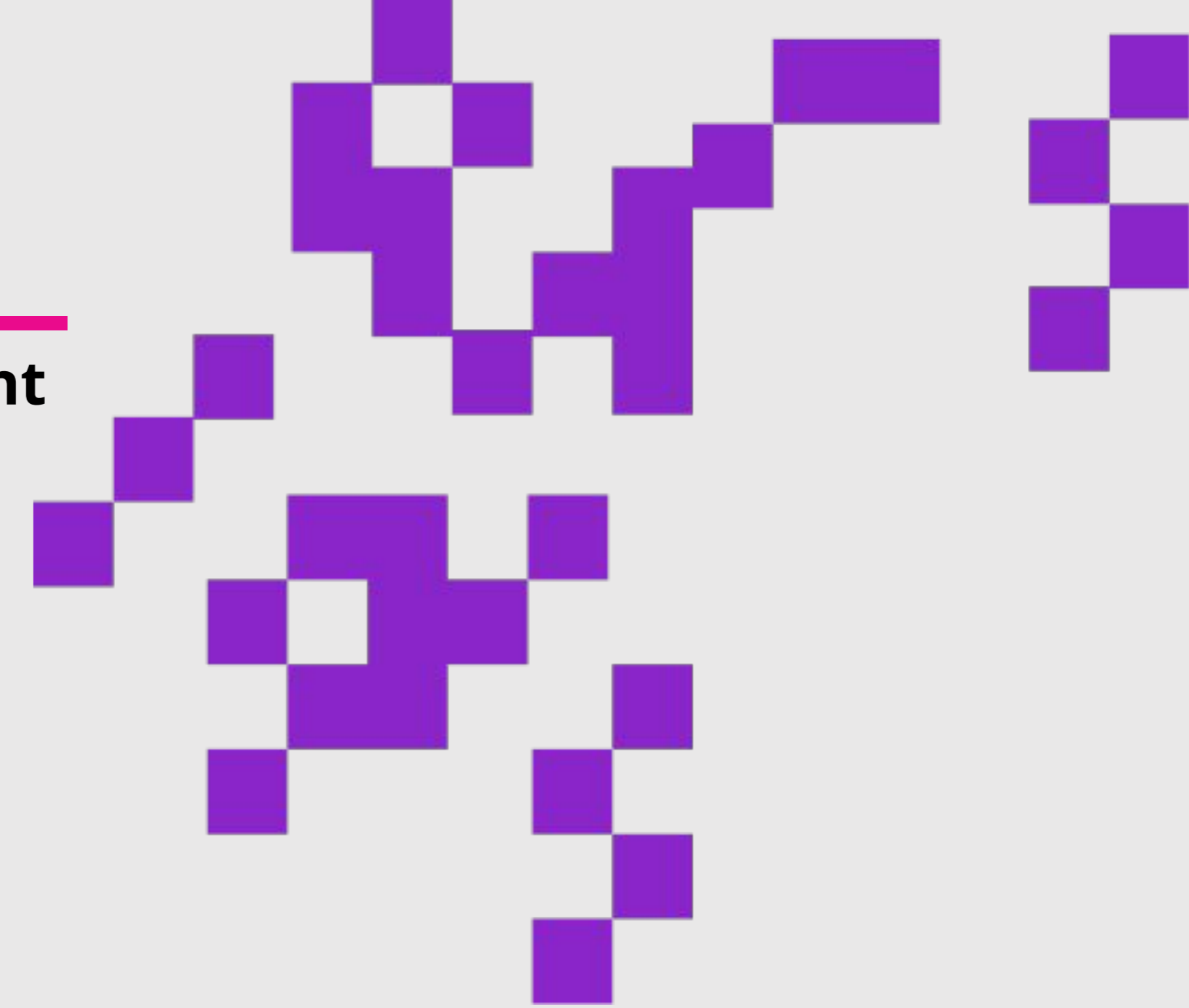
Set expectations about what your AI-powered product can do, cannot do, its risks, and how to improve it.

- Show users how to get the best results.
- Convey your product's limitations and risks in marketing materials and within the product so users can decide how they want to rely on the feature. Use hedging language when appropriate, e.g. "We think you'll like…".
- Let users know that their product experience will improve over time with their feedback. Inform users how, and ideally when, their feedback will benefit them. It will encourage future feedback submissions.
- Failure is inevitable and can't be avoided due to the probabilistic nature of AI. Inform users that they may receive errors, especially when they are new to your product, so that when they do they are understanding and are more likely to come back.
- Publish a terms of use that requires users to use the product responsibly and forbids use for violence or harm.

Have a plan for handling errors and failures so users can move forward with completing their task.

- Provide explanations.
- Provide user control.
- Provide customer support.

2. How transparent is your AI?

AI is not perfect, and you need to remind users of this. Avoid suggesting that your product is perfect and can fully replace a specific task, especially if your system's outputs are not yet reliable.

- Explain in general terms what your AI-powered product can do for your user and how it will benefit them. Give users just enough information to make a decision and move forward. The level of detail might depend on whether the user is a new user onboarding to your product (they might expect more detail), or a returning active user (they might prefer short explanations).
- Depending on your user, you can also explain how the AI model works if it will help users make smarter and more informed decisions about when to use it.
- Don't attempt to explain all the technical details within your product. You might not have the full understanding yourself, or the user could become so overwhelmed or bored that it leads to indecision or inaction.
- If you're proud of the technology powering your AI-powered products, or some users are asking for deeper explanations, consider sharing a technical blog post through your marketing channels or adding documentation to your help centre.

When helpful or necessary, **tell users how confident you are** in a specific prediction or recommendation, especially in high stakes situations, to help them gauge trust in the system and guide their decision-making.

- If possible, explain why the system provided a certain output, or why the system did not provide a certain output.
- Confidence can be communicated in a number of ways, including visual bar charts, percentages, rankings, attributions, or categorically (e.g. "best match" or "this price is likely to increase" or "because you read fantasy"). You will have to define the threshold for how to group recommendations, which may require user testing.
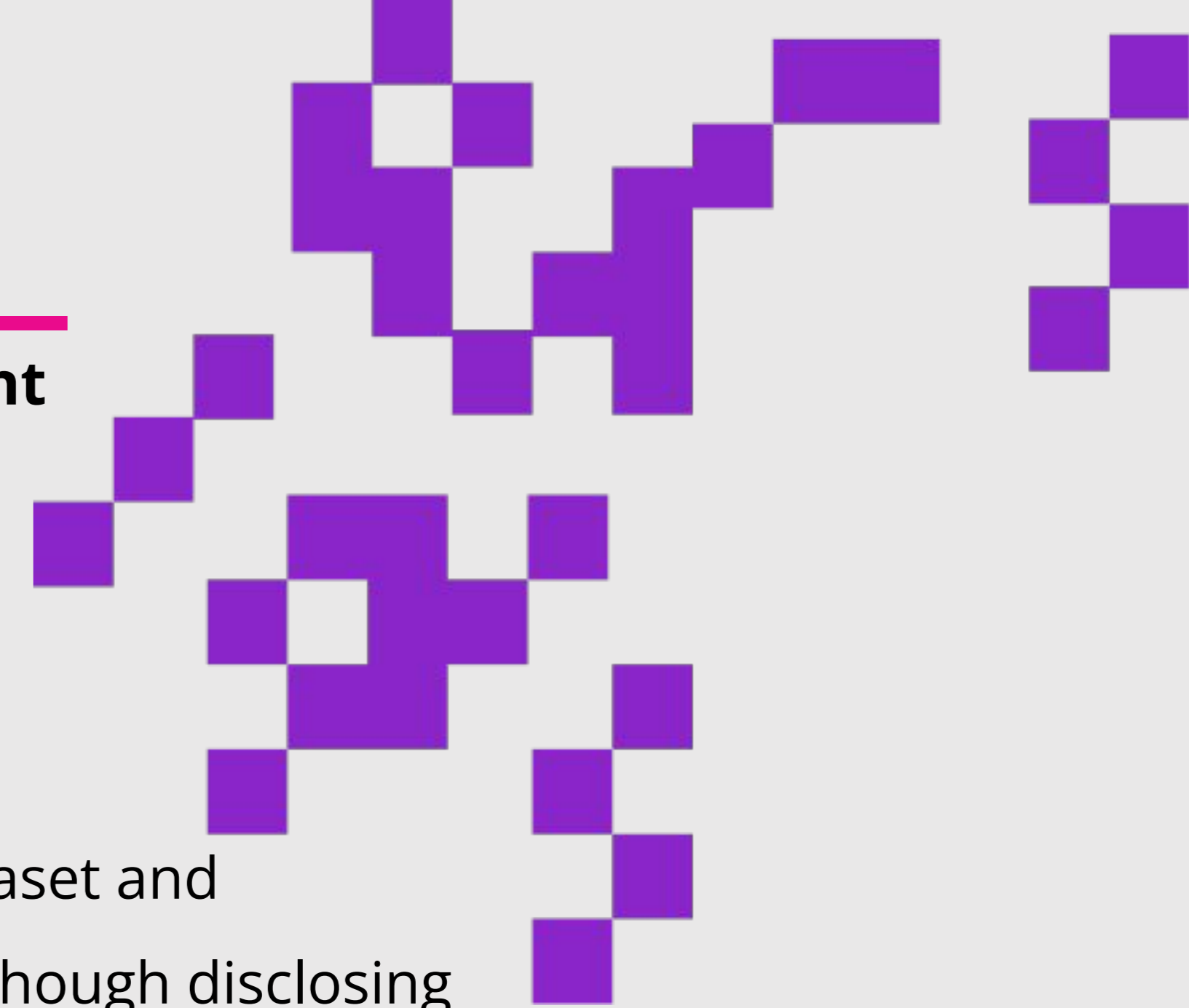
A released model should be accompanied by **documentation** detailing its dataset and performance metrics in aggregate and across different demographic groups. Though disclosing this information may put companies in an uncomfortable position, it is good practice. The documentation should capture the trade-offs across different metrics.

Consider **open-sourcing** your model to increase transparency and trust. It's important to note, however, that open-sourcing also involves potential risks, such as intellectual property concerns and the possibility of misuse. Carefully consider these factors before making a decision.

3. Do you give users options for personal control?

Give users the option to use the AI product or feature, if possible.

- Thoughtfully integrate the product into the user's existing workflow and make the use of its output voluntary.

Give users a way to give feedback.

- Provide access to human support through live chat, phone, or email.
- Allow users to submit feedback in-the-moment without having to change interfaces. For example, show a thumbs down emoji to rate a search result, allow users to skip a recommendation, or flag an output or a bug through a form.
- Refrain from requesting explicit feedback from users to avoid asking them to do extra work. Instead, use implicit feedback, such as favouriting, starring, or hearting to learn how people are using your product.
- Avoid asking users for both positive and negative feedback. Instead, just present them an opportunity to give negative feedback on predictions so they don't feel like they have to rate every positive result.
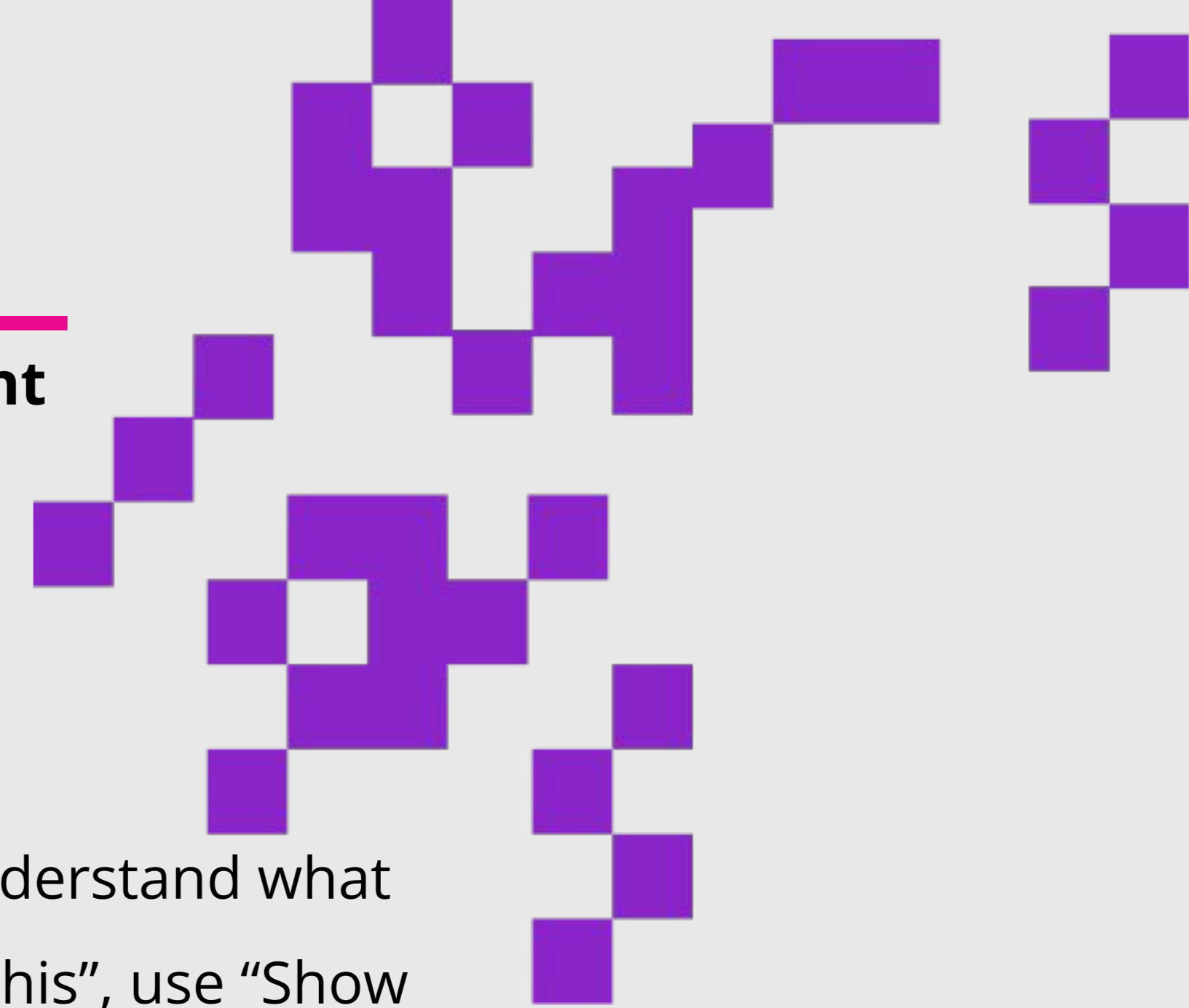
- Use simple and clear wording to describe feedback options so that users understand what happens when they choose the option. For example, instead of "I don't like this", use "Show me less sports news".
- If using icons, accompany them with a description to reduce ambiguity. Keep in mind that user feedback is subjective. A simple thumbs up or down can carry different meanings to a user, from "How is this thing so smart" to "This is good enough" to "I never want to see this again".

Give users a way to take over.

- When the system doesn't work or gives a poor prediction, ensure the user can still manually complete their task. This is especially necessary when first launching your product or feature.
- Allow users the ability to review, undo, or reject a suggestion and proceed with their own choice. If possible, give users immediate value or a reward for their extra effort.
- Give users autonomy and agency over their workflow so that they can develop comfort and confidence in this new way of accomplishing their task.

Give users control over their data.

- Ask for permission, don't plan to beg for forgiveness. It builds trust.
- Be transparent about the data you are using, how you are using it, how long it is kept, how it is secure, and give users an option to opt out.
- Limit the use of legal jargon and refrain from hiding data terms in fine print that requires scrolling and accepting to move forward. Instead, use simple and clear language in a friendly interface.
- Allow users to request information on their data that is collected.
- Give users the ability to adjust their privacy and data settings later on.
- Prompt users to review their preferences. People's preferences evolve over time, and sometimes people's preferences change after they have used something a few times.
- Give users a sense of what they can expect in output if they share data to help them decide how much they are willing to share.
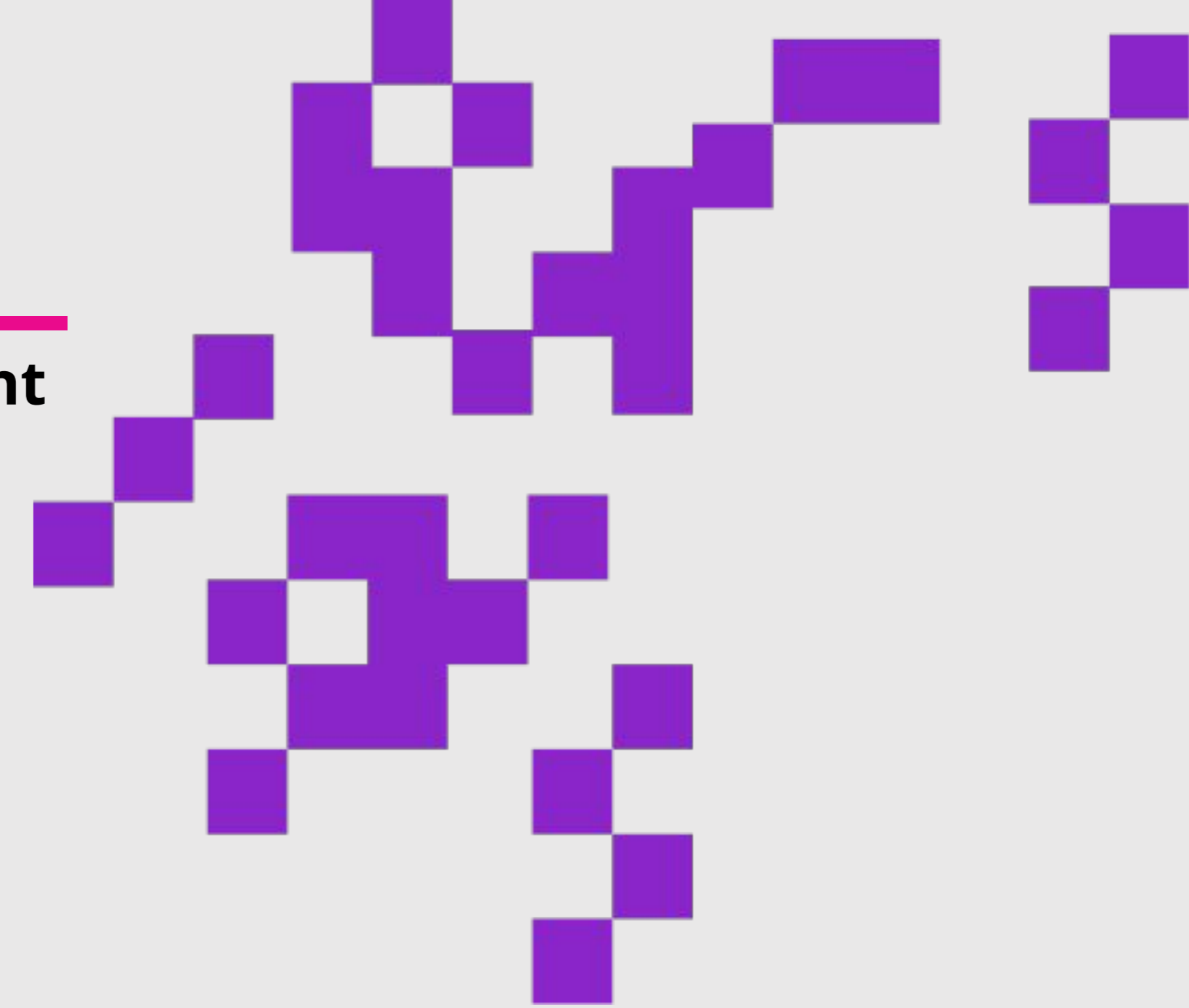
Post-deployment

What do you do once people start using your AI?

1. Are you building user trust?

2. Are you tracking user engagement?

3. Are you monitoring the performance of your model?
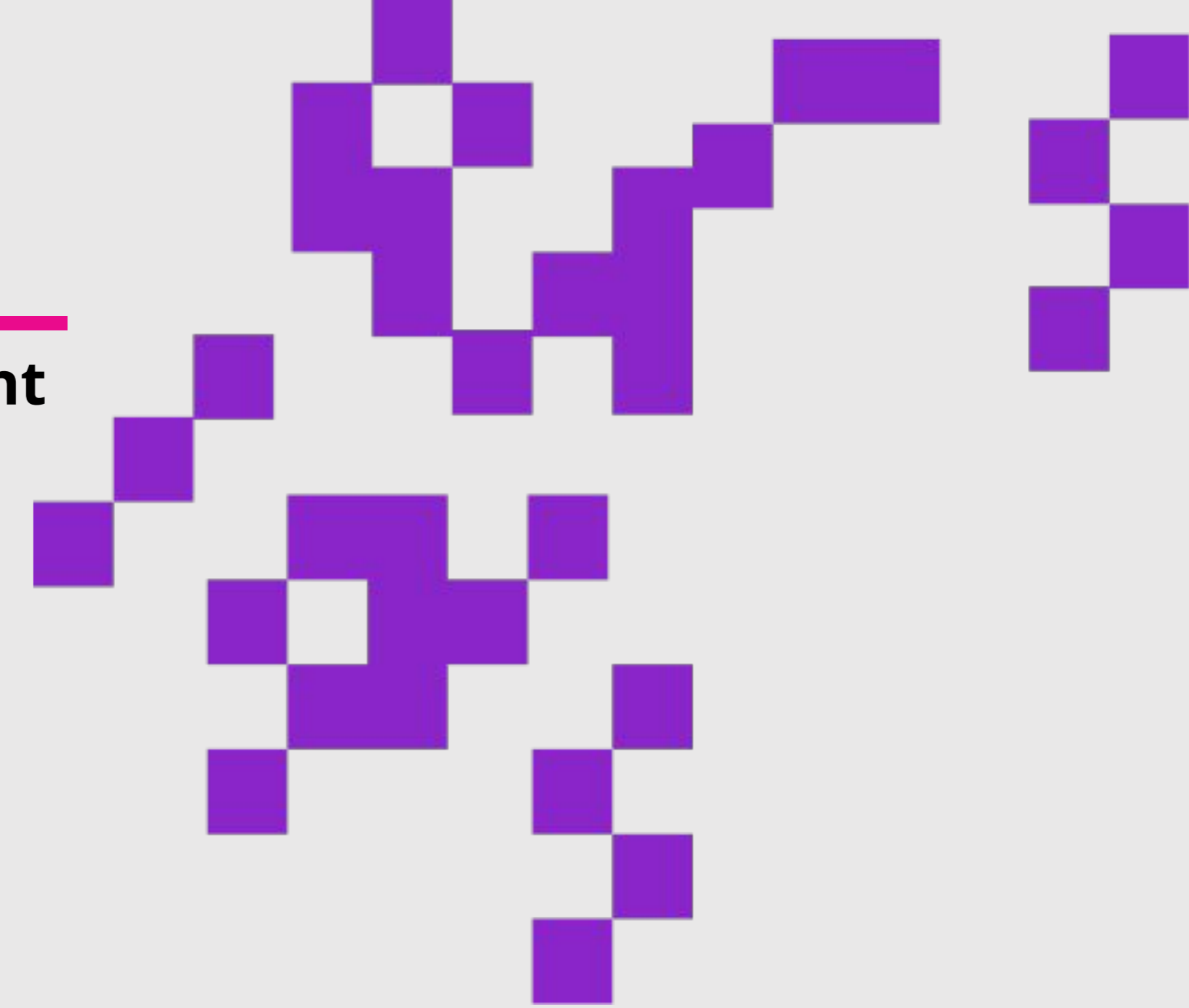
1. Are you building user trust?

Your model has been trained; your product is being used by real people. What now?

Maintain user trust.

- **Keep educating. Keep iterating. Keep transparent.**
- Inform users when you've made a significant change to your AI model or UX that might impact the predictions or recommendations they receive.
  - Active users might be more inclined to try AI improvements on functionalities they currently use.
  - Consider in-app announcements of noteworthy updates.

Revisit your feedback mechanisms.

- Prioritize recent feedback.
- Be cautious of confirmation bias. Implicit feedback mechanisms limit users to give feedback only on what they see, not what they might like to see or do.
- If a user's concern or problem cannot be addressed on the spot, share the feedback with the technical team and make changes to your product so that other users don't encounter the same problem. This is true for all software, but especially true for AI because feedback is integral to the system improving.

2. Are you tracking user engagement?

Observe how users are engaging with your AI-powered product.

1. Do a funnel analysis to understand how users are navigating defined paths within your product and identify potential problem areas where users tend to drop off.
2. Determine if there are certain conditions when a user is more likely to accept or reject a recommendation.
3. Observe trends and patterns over time and don't feel like you need to take action on every single piece of feedback you receive.
4. Look outside your core product for feedback, including Reddit, X, Facebook groups, or app store reviews.
5. Monitor for bad actors. Expect people to abuse your product, so regularly perform QA testing. It's a good idea to put failsafe mechanisms in place to discourage bad behaviour.
6. Evaluate whether your success metrics should change as users use your product more.
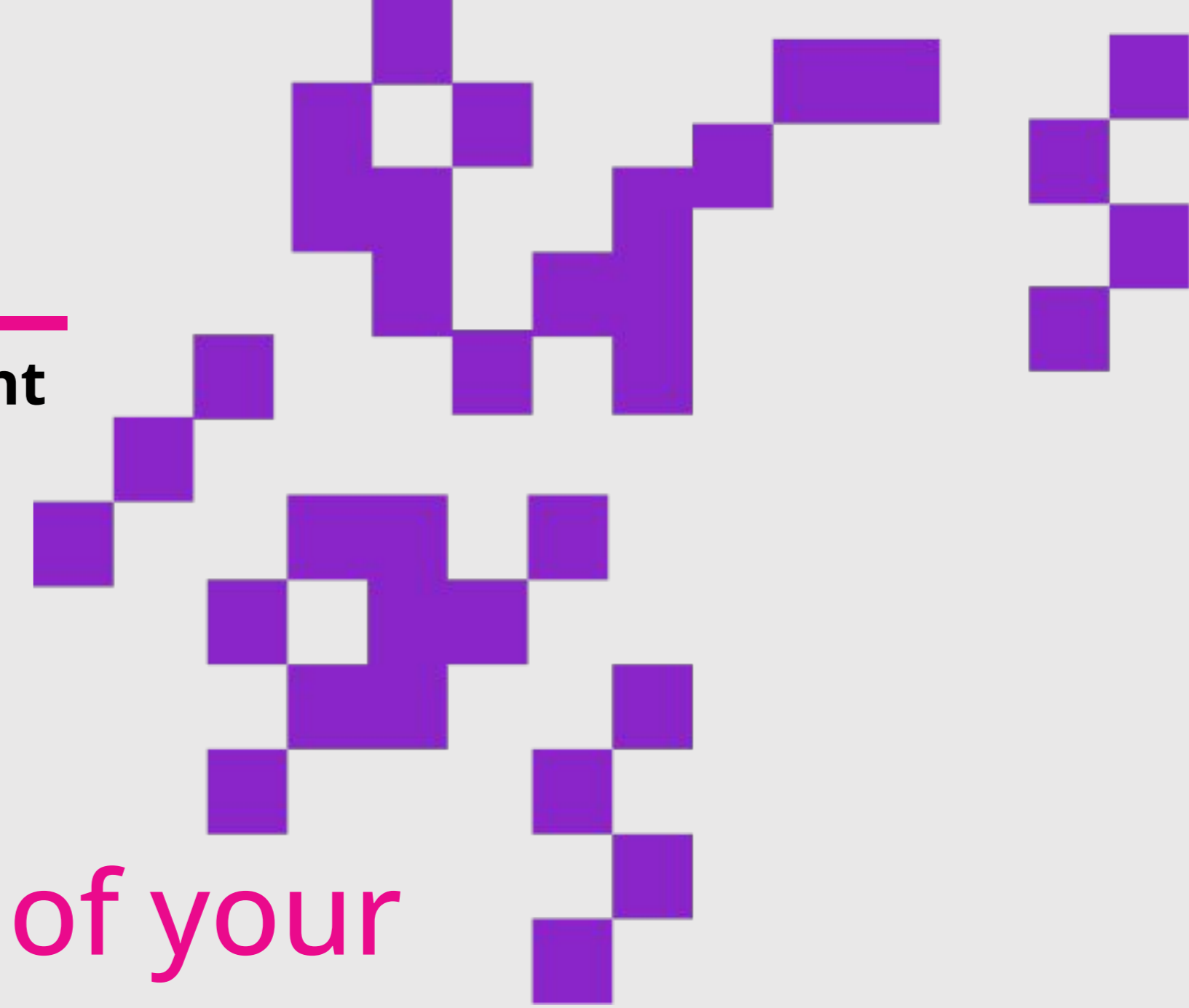
3. Are you monitoring the performance of your model?

Maintain your dataset.

- Start with the initial data quality analysis, but plan to maintain it over time.
- Compare your live data to your training data. If it's significantly different, you might be getting unexpected results and may need to retrain your model with new data. Retraining models frequently can be expensive, so consider the trade-offs between cost and performance.

Monitor your metrics on a regular basis and **set benchmarks** for certain actions that need to be taken when a metric drops below a certain threshold.

- Measure performance across user subgroups to assess fairness and harm.
- Audit your system for discrimination on protected grounds; consider a human rights audit if your product impacts historically marginalized groups or human well-being.
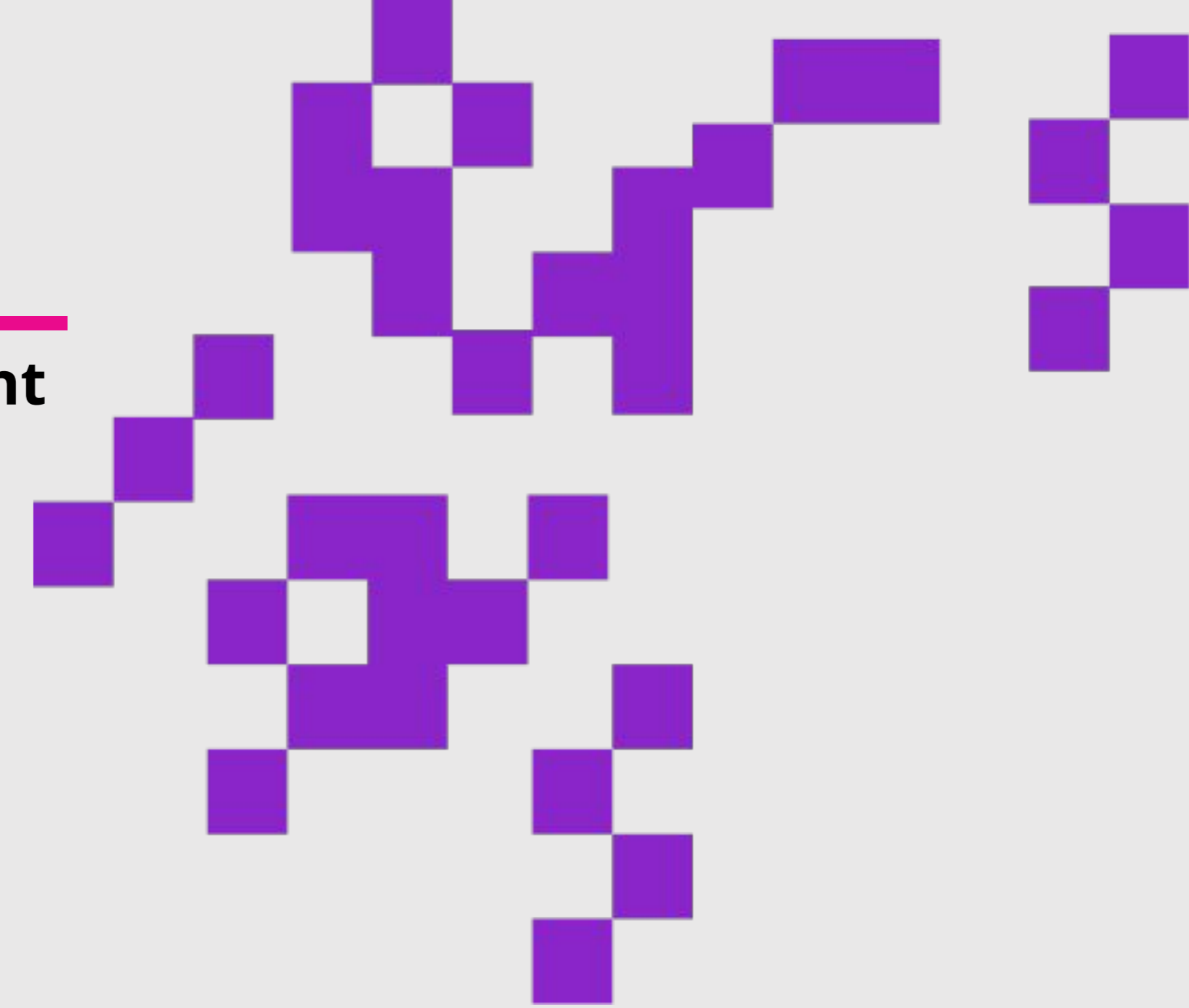
Assign someone responsible for monitoring the model after deployment and reporting performance.

Share model performance reports, detailing the performance and improvements made, with stakeholders.

- Use visualizations and simple indicators to help non-technical stakeholders understand the system's performance.

Understand that iterations to your model take time; machine learning engineers and data scientists will want to run experiments, wait to deploy until the next version release, or get more data.

Stay updated on new regulations and the dynamic AI landscape to maintain compliance and adapt to evolving legal and ethical standards.

Thank you

The Principles in Action Playbook incorporates insights from Vector Institute user interviews, alongside existing academic and industry research.

- Amershi, S., et al. (2019). Guidelines for Human-AI Interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.
- Design for AI. IBM. (2022). https://www.ibm.com/design/ai/
- Llama Responsible Use Guide. Meta. (2024). https://ai.meta.com/static-resource/responsible-use-guide/
- Machine Learning. Apple Developer. (2023). developer.apple.com/design/human-interface-guidelines/machine-learning
- People + AI Guidebook. Google PAIR. (2021). pair.withgoogle.com/guidebook
- Pistilli, G., et al. (2023). Stronger Together: on the Articulation of Ethical Charters, Legal Tools, and Technical Documentation in ML. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency.
- Rules of Machine Learning. Google. (2024). developers.google.com/machine-learning/guides/rules-of-ml
- Shen, H., et al. (2021). Value cards: An educational toolkit for teaching social impacts of machine learning through deliberation. Proceedings of the 2021 ACM conference on fairness, accountability, and transparency.
- Varanasi, R.A., Goyal N. (2023). "It is currently hodgepodge": Examining AI/ML Practitioners' Challenges during Co-production of Responsible AI Values." Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.
- What is Responsible AI? IBM. (2024). ibm.com/topics/responsible-ai
- Yildirim, N., et al. (2023). Investigating How Practitioners Use Human-AI Guidelines: A Case Study on the People+ AI Guidebook. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.

We'd love to hear from you. Please get in touch!

About Vector Institute

Launched in 2017, the Vector Institute works with industry, institutions, startups, and governments to build AI talent and drive research excellence in AI to develop and sustain AI-based innovation to foster economic growth and improve the lives of Canadians. Vector aims to advance AI research, increase adoption in industry and health through programs for talent, commercialization, and application, and lead Canada towards the responsible use of AI. Programs for industry, led by top AI practitioners, offer foundations for applications in products and processes, company-specific guidance, training for professionals, and connections to workforce-ready talent. Vector is funded by the Province of Ontario, the Government of Canada through the Pan-Canadian AI Strategy, and leading industry sponsors from across multiple sectors of Canadian Industry. For further information or media enquiries, please contact: media@vectorinstitute.ai