Navigating the Ethics of InvokeAI: A Responsible AI Assessment

Parker Christenson, Gabriel Colón, Dominic Fanucchi

Shiley-Marcos School of Engineering, University of San Diego

AAI-531-01: Ethics in Artificial Intelligence

Maryam Keshtzari, Ph.D.

April 8, 2024

1.0 System Description

AI/S Purpose and Description

InvokeAI

Authors: Parker Christenson Dominic Fanucchi Gabriel Colón

Institution: University of San Diego

Peer Reviewers: Steve Amancha, Ahmed Ahmed

System profile

System name	InvokeAI
Team name	Team 7

System description

Our group is looking into the generation of Artificially created images given a text prompt. Invoke-AI is an open source system, as we want to be able to investigate the models creation, training, and methodology. We realize that open source models allow for other developers to use the model in their own creations, free of charge. Considering that the machine learning model could be used in other applications, we wanted to evaluate the ethical and moral concerns, if what the machine learning model is creating should be considered art, and if the art is infringing on other artists' style. (Fairlearn, n.d.) We think this is a very important thing to talk about and consider because of implications of artificial intelligence being trained to replicate artists' work and or style, with the consideration of potential copyright infringement.

Description of supplementary information	**Link**
Github page	https://github.com/invoke-ai/InvokeAI/
Links to both the training set and validation set	Validation Data Set and Training Data set (Github .p files)
InvokeAI's About page	https://www.invoke.com/about

System Features

Features	Description
Natural Language Processor	Processes text input provided by the user based on a training data set to understand the words, phrases and sentences.
Computer Vision	Model is trained on hundreds of thousands of images
Stable Diffusion	Deep Learning model for converting text to images
Hugging Face	The platform where they save their model and expand on their training and validation data sets.

System Purpose

AI Actors	AI Users	AI Stakeholders
• The AI actor in this case depends on the fact that the people who want to adjust the model and add features, can do so and make it their own. The way that the model sits right now, the AI actor is Invoke AI.	• The users of the model are the average person, who interacts or gives an input into the prompt, if that is a 3rd party system, or directly from websites (Fairlearn, n.d.). • Companies that decide to utilize the open source software to develop products	• The stakeholders, are companies that choose to use Art generation via, text to prompt or any sort of method to communicate in language to the model • Art/Creative Communities creating the original art upon which the model is being trained on • Financial Investors

Deployment mode

How is the system currently deployed?	The system is currently deployed via an API that the user can use at various websites, or they can download the model directly from Github to be able to interact with the model locally on their own machine.
Are there different deployments?	As of right now there are only two, via an API or via a local instance.

Intended uses/features

Name of intended use(s)	Description of intended use(s)
1. Inpainting/Image Editing	Inpainting is a process of filling in missing/masked portions of an image using the stable diffusion model. This could be used by any users that are looking to remove unwanted elements from images, restore damaged images, and modify images to the users desire. This system feature can be accessed using the web UI or CLI.
2. Image generation	Using text prompts, the system is able to generate high-quality images using stable diffusion models. This feature is intended to be used by artists, designers and even AI researchers. This feature can be accessed using the web UI or CLI.
3. Upscaling/Image enhancement	The system includes tools that allow for upscaling and enhancing generated images. This would be particularly useful for artists/designers that need to create high-resolution images/graphics from low-resolution images. This system feature can be accessed using the web UI or CLI.

2.0 Summary: Ethical Risks and Impacts

Ethical Principles of Artificial Intelligence

AI Principle	Definition, related principles, application to the sector
Trustworthiness	Artists may not trust the technology if their work is constantly being infringed upon.
Dignity	The AI model should not generate images that would be publicly inappropriate, such as pornography, especially using the styles of artists.
Freedom	Generated images should be considered public domain, and should be available as examples of the AI's artificial imagination.
Fairness	The compensation of the artist's work should be proportionate to the use of their work, especially if the resulting image is especially requested to use that artist's style.
Bias	Algorithmic discrimination.
Privacy	Generated images should not be traced back to the account of the prompt used to generate the image, thereby maintaining user privacy.
Safety	Images being generated should not be inciting violence or to threaten another person.
Transparency	Image Generative AI companies should provide the data they used to train the model.
Explainability	Image Generative AI companies should be able to explain their processes and how the model works.
Accountability	If the data used to train a model is using the artwork of the artist without their permission and/or compensation, the company should be held liable.
Governance	Policies regarding copyright infringement should be applied on behalf of the artist

Overview of Ethical Risks and Impacts

AI potential harm	Why it can happen
Stealing Artists trademarked Styles	It can happen because Large language models do not fully understand the capacity of what they are doing. LLM's are only able to understand the text they are fed, and unless they are specifically told that they can not respond to specific texts/prompts, then they will respond. Chat GPT-3 was able to provide credit card numbers because the user prompted Luhn's algorithm, until it was later patched, the same idea can be applied to art. The end user can give a prompt to paint or make an image similar to someone's art by providing examples (Knibbs, 2024), the model produces it, and the end user could pass the art off with the intent to make a monetary gain.
Invoke AI is not fully transparent, even though their code is	While Invoke AI's code might be open source, we are able to see the 'full amount' of images the model was trained on. This is a concern because they used a 'bootstrap' like technique, and they did not produce or label a full dataset. On the other hand, if they went through 50,000 images, and labeled them all themselves, that they acquired web scraping, then there is no issue to be had. But the bootstrapping technique does not actually produce images, but rather builds off of other images using the hugging face library(Fairlearn, n.d.).
Who will govern the machine learning models, and how will law be applied to the companies that create them	There is not really a way to govern these kinds of things as of right now. There have been some talks of AI generated music, and right now there is a lot of confusion on if the AI generated voice is in fact breaking copyright, because the AI is technically using the artists voice, rather than stealing lines or musically trademarked beats. We think that there will be a time, where and Artists like Jackson Pollock's art gets re-created, and will that be copyright infringement?
Safety of the general population	There are currently no regulations on allowing AI to produce explicit or illegal content. This is a huge problem, because even though the images are not real, the acts that they portray can be considered realistic enough. In 2023 there was a market crash because a fake photo of the white house was published that made it seem like it was attacked. This cost real companies and people money, and to our knowledge, there was no legal action for the individual who prompted the photo to the machine learning model. This could very well cause harm to people in the feature, for various other reasons.

Accountability of the thing who harmed the individual	Who will take accountability when these images cause people serious harm? With people advocating for AI and image generation not to be regulated, and we think that there should be someone held accountable. As of right now, there are very few laws that will hold people accountable in these situations.

Ethical AI Principles: Ethical Risks and Impacts

1. Use the module resources to define the goals or requirements of the ethics principle (e.g., Principled AI report).
2. Complete Stakeholder Impacts iterative analysis templates (Tasks 1-3).
3. Update the summary table with an indication of harms/risks the principle in focus.

Freedom

Goals/Requirements	Ethical Risks
Freedom to use public domain images to train models without requiring legal permission	Risk of using an image that copies an artist who has not given consent to have his work used (Hariri et al., 2019).
Freedom to use AI-Generated for self interests	Self interests of users may not be align with ethical values
Freedom of expression in AI-Generated Art	Risk of creating an image that may be inappropriate for some users.
Freedom to take down any generated image that violates ethical values.	May violate a person’s right to express themselves.
Freedom to express discontent with the model's performance.	Would decrease public opinion if the company is not willing to take feedback.

Dignity

Goals/Requirements	Ethical Risks
Make the AI system Generate Safe Images	This will be hard but there should be a detection system that is able to detect Explicit requests
Do not allow for very niche styles to be promoted	This will again also need to be limited in the sense of making sure prompts are restricted or key words to ensure that specific styles aren’t stolen
If the AI is generating Hyper Realistic photos, add watermarks	This ensures that the images that are being produced aren't actually going to be used to cause harm (Knibbs, 2024)
Keep the consumers trust at all time	make sure that you make certain parameters about your model well know, even the ones people find hard to understand
If there are errors with the model, be transparent, and own up to the errors	The ML model is almost never going to be perfect, but the people who back the model can take the fault whenever there is an error with it.

Fairness

Goals/Requirements	Ethical Risks
Depict historical figures appropriately	Failure to depict historical figures accurately could be deemed as revisionist history, and paint the company in bad light.
Train the model with images that represent all cultures in a fair light.	If the AI-Generated images are insensitive to cultures this would harm the company’s reputation.
Train a model without violating copyright laws.	Copyright infringement on an artist's work is unethical and would result in fines for the company.
Models should not favor one political view over another when generating images about politics.	Political bias would paint the company as supportive of a given political party and alienate others from using the model.
Model should not be discriminatory against religious beliefs, or protected classes.	May infringe on the freedom of speech/expression.

Bias

Goals/Requirements	Ethical Risks
Make sure the AI does not create images favoring one side automatically	This will cause uproar within the political think-sphere as the model will be 'Biased towards one side' and we want the model to not favor anything/anyone
Make sure biases can be added through prompt	Allow the ML model to be prompted to have a bias but, ensure that the model does not come with one from factory
The data set should be trained on the best case fit, and then bias be evaluated	The model should be trained to have lots of images in the dataset however, the model should not be limited because they want an even split of all images to claim it was trained without bias
The model should be audited prior to be released	The model should be looked over by a third party, and the third party should be able to look over all aspects of the model prior to deployment.
Bias should be announced if there is anything	If for some reason all fails, and the stakeholders want to release a model that is imperfect, the stakeholders should come out and announce the flaws prior to the release, and how they are trying to fix it with a patch.

Security/Safety

Goals/Requirements	Ethical Risks
Protect user privacy and prevent exploitation of personal data in model training	No consent to opt-in regarding training data
Prevent misuse of AI art generation	Production of deep fakes; risks of unauthorized access
Ensure AI art systems are secure and resilient against hacking	inadequate cyber security measures
Promote safe deployment and lifecycle management of AI art models	dangers of outdated generative models
Detect mode Drift	Model drift should be used to ensure that the model is not giving inaccurate predictions and that could potentially be harmful to some users.

Privacy

Goals/Requirements	Ethical Risks
Protecting the privacy and personal data of individuals involved in the AI art generation process	insecure handling of sensitive data (biometrics, facial images)
Obtaining consent from people before using their data for training models	lack of clear consent mechanisms and opt-in processes
Upholding people's right to privacy; prevent unauthorized generation of sensitive content	inadequate techniques to block generation of explicit material; prevent generation of non-consensual images
Ensuring and providing transparency surrounding AI art technology	disclosure of when outputs contain personal data or likeness of individuals used for training
Limit Web Scraping	Limit web scraping data collection methods, as the people that you are downloading the images from, do not necessarily consent to having the image taken.

Trustworthiness

Goals/Requirements	Ethical Risks
Mitigating risks of misuse, bias that would undermine trust of the system.	Inadequate techniques to detect and prevent generation of harmful, explicit and potentially illegal images.
Creating and fostering trust from the public and instill confidence surrounding the development and use of AI art systems.	Failing to demonstrate a commitment of ethical principles.
Explaining the capabilities and limitations of AI art systems.	Inadequate information and disclaimers about the generated outputs.
Develop and maintain reliable AI art systems that will consistently produce high quality outputs.	Inadequate quality control measures.
Validating and testing AI art systems before deployment	Risks of deploying an AI system without sufficient evidence of its trustworthiness.

Explainability

Goals/Requirements	Ethical Risks
Model should output images that make sense.	If the images generated don't make sense the credibility of the model and company would decrease, and the risk of something unethical being generated on accident would exist.
The general processes used by the model to generate images should make sense to the public at large.	Lack of transparency with the public on processes may create a bad reputation for the company concerning its internal workings, especially if there are suspicions about copyright infringement or unethical properties of the model.
Model is required to cite artists whose work was used to train on.	If an artist's style is used in an AI generated image without citing the artist, the company could risk being sued by the artist.
The model should be able to explain how it came up with an image.	If the model cannot explain how and why it generated the image the way it did, it could result in lack of trust from society.
The model should provide an explanation if it doesn't generate an image from a prompt request that violates company policy/ ethical principles.	Lack of information or explanation when a prompt for an image request violates company policy or ethics, can result in decreased trust, reputation, and could fail to inhibit attempts to generate unethical images.

Transparency

Goals/Requirements	Ethical Risks
Make the model as open source as possible	Some people will modify the model to not function in the way of the founders intention
Have the model be open source	Let people contribute to the project, and even have people offer to use their Artwork in the training materials
Have a well document Github	This will allow people to investigate the model and see what is being done in the source code, but will allow others to change it.
Show the training errors	This lets people know what the model is bad at doing, but people might be able to take advantage of the wrongdoings, depending on what they are.
The founders should be transparent about their motives with the model	The founders of the model should be open, and say their intentions with the model, and if their goal is to replace aspects of society, then they should come out and say that.

Accountability

Goals/Requirements	Ethical Risks
Create frameworks that clearly define responsibilities and accountability for companies creating these systems	lack of defined roles, responsibilities and accountability
Create systems for those hurt/damaged by the output of AI systems	lack of defined channels for grievances and processes for investigations of complaints
Create systems for performing audits	Technical limitations for logging, documentation and audit trail capabilities
Have the founders own the whole process	Ensure that the head ML engineer and the founders are taking full ownership of what they put into production, this ensures the model will be up to their standards, and won't be a PR nightmare.
Make sure the model is able to be removed from deployment rapidly	Make sure that the model's deployment method can easily be stopped, if there are issues that arise from potential biases.

3.0 Summary: Ethical Feasibility

Legal, social, and environmental feasibility

	Legal	Social	Environmental
Description	Requiring permission from artists to use their work for training models.	distrust in the system and the technology surrounding it due to using works of artists without permission.	e-waste from hardware and more demand on the power grid.
Question 1	Q: Is there a process in place to get permission from the artist to use their work (legal paperwork)? A: There is currently not general forms being used by Invoke AI to allow people to submit their art	Q: Is the culture surrounding art impacted by the AI/S system and process? A: The AI culture is very two sided, one side is for AI generated art, and the other is against AI generated art, and more for a purist artform.	Q: Where does the e-waste go after storing the models data for training? A: There are some companies that have been reselling used server parts, all in order to reduce E-waste.
Question 2	Q: Is there a process to check for unauthorized use of artist work (copyright issue)? A: There are currently no processes that are available to check for plagiarized art; however, there are AI models that detect text plagiarism.	Q: How would the arts market be affected by the influx of newly generated AI art? And would human created art be valued more? A: AI generated art would flood the market and be used in very general settings. Human generated art would be specialized and increase in value.	Q: What happens to the datacenter hardware after the GPU's lifecycle is over? A: Hardware is sent to a recycling center where it is taken apart and prepared for the remanufacturing process.
Question 3	Q: Are there penalties in place for violation of trust by companies? (policies, rules, laws) Q: There have been recent efforts from the Biden Admin to push for some controls on the AI growth in order to mitigate harm.	Q: Would people be scared of having data/images harvested without their consent? A: The general idea is that the public would have to consent somewhere in the EULA, that states that their data could be harvested.	Q: What powergrid changes need to be made for all of the data centers that open? A: Solar Panel systems will be installed on the roofs of all data centers. Other green energy sources should also be considered.

Question 4	Q: Should companies be required to have a published set of values that encompass ethical practices with data? A: There are some companies like Rolls-Royce, that are very upfront about their ethical processes.	Every community is different, what are some concerns that can not be addressed by just a blanket policy? A: There are some concerns about people's culture not being preserved anymore, and the tracing of the culture will be lost in the Data age.	Q: What environmental impacts does the increase in electricity have? A: The impacts will be higher at first, and it's therefore proposed that we collaborate with other AI companies to build communal green energy sources and invest in future technologies that will surpass demands.
Question 5	Q: What penalties should exist for companies that infringe on copyrights? A: Currently US lawmakers are pushing for more regulations on the development of AI.	Q: How would the impact of having ML generated art impact people's livelihood who rely on selling artwork/art pieces? A: People who have been making art and hyper-realistic images, have started to notice that their specific job market has started to decline. (Carter, 2023)	Q: What happens if the power grid can not support more data centers? Who pays for that powergrid update? A: This is where the collaboration with other AI companies will be useful. We need to come together and raise the funds for the power grid upgrade. This will require a funding plan.
Question 6	Q: How will intellectual property rights be managed for AI-generated art? A: That will be a hard one to predict due the current state of the political climate, there could be a time where the model owner owns the art, and the person's image that borrowed inspiration from, wont need to be credited.	Q: How would AI-generated art impact the artistic development of new generations? A: Art development could go two ways. More art will be developed, but there will be a large decline in the hand made arts, and the development of the hand done art and styles.	Q: What is the long term effect of having more than one data center up and running? What type of power plants would be made? (Coal, Natural gas, Nuclear, wind, etc…) A: Nuclear power is the cleanest source of power and should be the energy source that we would ask our government officials to consider. Wind and solar are good options as well.

Highlight factors and conditions that make the AI/S ethically unfeasible, present significant risk to individuals and society under current deployment, near-term or future.

Current Deployment

- There are currently no general forms being used by Invoke AI to allow people to submit their art or by any company for that matter

This brings up some ethical concerns, as the art that is being generated by these models, are technically not creating their own style but rather referencing the art made by others.

- Current models are made safely by submitting prompts to an API

This method of deployment can very easily be turned off. Machine learning models are deployed using API's because they are effective to implement into a webpage but the added benefit of being able to shut down the model from a remote location from nowhere near the host server, is also an added advantage to the API deployment method.

Near-Term

- Certain laws regarding copyright infringement could be ruled on shortly from AI generated music which could affect AI generated art.

The music world has also been taken by storm as there are AI models that have been able to replicate peoples voices. The artists that are the subject of these models stealing their voices have filed lawsuits against these companies. Their courtroom decision could very well be precedent in some artists' cases. The precedent in those cases will make way for all other AI generated art rulings.

Future

- Implementing watermarking on images that are generated by AI in order to prevent copyright infringement, and potentially dangerous AI generated images.

There will be a need to have a type of watermark for AI generated art to distinguish them from human generated art and a way to cite artists whose art was used as inspiration for the AI generated art. This could be some type of QR code that takes viewers to a list of artists and styles used.

4.0 Practices: Mitigation of Harms

Non-Technical Risks and Harms

Summary of critical risks and stakeholders related to non-technical issues

Type of Harm	Stakeholder	Possible tools/strategies for mitigation
Copyright infringement	The company, and or the investors	Prompt limitation - Prompt limiting is the idea that you are able to limit certain strings of text, to mitigate the outputs of the model.
Economic loss	The people who produce art	Limiting the kinds of artistic styles that are in the model training data- Limiting the models ability to make certain images by restricting the training size.

Technical Risks and Harms

Summary of critical risks and stakeholders related to technical issues

Type of Harm	Stakeholder	Possible tools/strategies for mitigation
Lack of Transparency	Users,	**AI Now Institute Algorithmic Accountability Policy Toolkit -**
Bias model/lack of fairness	Developers, Users, Investors	**Agarwal et al (2018) A Reductions approach to fair classification -**

Conclusions: Mitigation

Stakeholder Engagement

Action Items (Non-Technical)

Summary: List Actions and Solutions to Follow-Up

Action Items:

Limit the training data - Making sure that the training data is limited to only user submitted art. The goal is not to webscape data, and collect millions of images (Knibbs, 2024), but rather create a curated style for the model rather than creating a generalized model trained on hundreds of art styles.

Control the prompts to exclude copyrighted art- Make sure that art that is under copyright cannot be replicated. With that being said, there cannot be a perfect first model but, rather the developers must continue to limit the prompts and make sure that the art that is in fact copyright, becomes more and more difficult to replicate.

Make sure the economic loss is mitigated- Within the transfer from human held jobs, there will be a human economic loss, on the contrary there should be some kind of economic revenue gain from other kinds of means. For example, if people lose jobs the GDP should not go down, but the model should create new taxable revenue streams to keep the GDP relatively the same.

Make sure that the model is made transparently- The AI stakeholders should show that the model is in fact made with all of the images, and not bootstrapped. Using libraries like Hugging Face makes the model more susceptible to creating images that are not directly asked for and increases the risks of the company to be under more scrutiny from the general public and the government for various law infractions (Norwest Venture Partners, 2024).

Gain the public's trust with the model- Make sure the model is able to do tasks that are commonly asked for if the model is going to be released to the public for their own personal use. The model should be able to use and produce the provided resources, and provide the best possible outcomes, as close to the end users request as possible. The model should be able to look at the input which more often than not, will be text and decide quickly, accurately and safely if the image should be produced. The model might allow for image morphing, but the model should understand that the prompt being given is allowed, or not allowed.

Roadblocks to Ethical AI (Non-Technical Solutions)

Action Items:

1. Political counterparts could still group ethical AI with the economic effects of AI in general, that is taking away jobs from their constituents.
2. There can be fear from a socio-political perspective that artists would lose their value in society, we may become dependent on AI for the creation of art.
3. Create copyright laws and regulations that specifically target the Generative AI industry and promote the building of ethical AI.
4. There are currently no laws or frameworks that allow or prevent the model to produce pornographic images of people without their consent. The production and distribution of revenge pornography is illegal but the model should realize or identify the changes that they are about to make to the image, could potentially break laws.
5. There are currently no laws regarding the way that the language needs to be ‘brightened or enlarged’. There are currently very little to no laws around the world that will protect people from malicious attacks, or copyright infringement.
6. If AI enabled technology is deployed and expands without the implementation of our recommendations, there will be a high likelihood of harm to individual reputations, and lack of trust from society on Generative AI Companies.

5.0 Tools: AI Ethics Toolkit

Ethics Toolkit

Ethics Tool	Target Risk/Harm	Stakeholder	AI Process Stage	AI Process Stakeholder	Source	Degree of relevance to principles							Consequences	
	Intended/Unintended outcomes or impacts to mitigate	Affected User Stakeholders	AI Lifecycle ML Process Step Source of Harm Solution Design	Ai Actors, Investors Project Team, Enablers, Government Organizations	Related links Examples Research/Community	Dignity Human Rights	Bias Fairness	Privacy	Safety Security	Traspa. and explain.	Accountability	Trust	Benefits	Drawbacks
New Economic Impact Model	The tool kit evaluates the following, considering their economic outcomes, Positive- Job creation negative-job displacement	The tool kit evaluates the following, considering their economic outcomes, artists, businesses, consumers, Developers	The tool kit evaluates t, their economic outcomes, and if the model will cause harm to the source of the data	The tool considers if the model will affect any of the mentioned economically and how the outcomes could end up	http://ethicskit.org/downloads/economy-impact-model.pdf	Low	Med	Low	Med	High	High	High	The benefits include the model consideration for remove jobs, and if the economic impact is to great, than it shouldnt be built	The drawbacks in the tool, also don't want the model to be produced if the economy can't support it, or if it's publicly funded.
a3i the Trust in AI Framework	Targets security risks, and lack of transparency. Positive- public trust negative- costs	Users	Deployment	Data Security Team		Low	Low	High	High	High	Med	High	Designing a system, with safety, security, and explainability in mind. Trust from the public.	increase cost to maintain security of data, infrastructure and model.
Agarwal et al (2018) A Reductions approach to fair classification	lack of fairness Positive- fair model negative- oversimplification of fairness	Users	Deployment/ model training	Development Team	https://github.com/Microsoft/fairlearn	High	High	Low	Low	Low	Low	Med	Generated images have fair representations of cultures, peoples, ect.	May impact production deadlines
AI Now Institute Algorithmic Accountability Policy Toolkit	Lack of Transparency, violation of laws/policies Positive- Trust negative- increase cost	Development team, investors	Model training	Development team	https://ainowinstitute.org/aap-toolkit.pdf	High	high	Med	med	high	high	med	Makes sure laws and policies are not being violated	Increases time to develop algorithm, directly impacting costs
Hesketh. Ethics Cards	Not considering negative impact the AI idea may have Positive- Start out thinking with the end in mind, and	investors. project team,, developers	Business Development	project team, developers	http://ethicskit.org/downloads/ethicskit-cards.pdf	high	high	high	high	high	high	medium	Start by proactively thinking of the impact the model may have and looking	increases the time to prepare for development.

	issues negative-increase time to launch												for solutions	
Kroll et al (2017) Accountable Algorithms	lack of transparency in code Positive-feedback negative-code becomes public	developers	building	developers		low	low	low	low	high	med	high	public perception of the company would tend to be positive (nothing to hide), feedback would help improve the model	competitors can use our code.
Makri and Lambrinoudakis (2015) Privacy Principles: Towards a common privacy audit methodology	lack of respect for user privacy Positive-public trust negative-increase time for several processes	developers, users	data procurement, monitoring	developers		med	low	high	high	high	high	high	increase trust from the public, and make sure there are no privacy violations.	Increase cost by adding privacy auditing, and directly incentivizes for data security team
Mitchell et al (2019). Model Cards for Model Reporting	lack of model use case scenario research Positive-assist in model development negative-	developers	building	developers		low	low	low	low	low	low	low	Provides examples for which models to use for what. May help choose a model fasters and start building	If it's the only resource used for model selection research, we could lose on learning about other perspectives.

References

Bender, E. M., & Friedman, B. (2018). Data Statements for Natural Language Processing: toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, *6*, 587–604. https://doi.org/10.1162/tacl_a_00041

Carter, T. (2023, October 3). Workers are worried about AI taking their jobs. Artists say it's already happening. *Business Insider*. https://www.businessinsider.com/ai-taking-jobs-fears-artists-say-already-happening-2023-10

Ethics of artificial intelligence. (2024, February 8). UNESCO. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

Fairlearn. (n.d.). *GitHub - fairlearn/fairlearn: A Python package to assess and improve fairness of machine learning models.* GitHub. https://github.com/fairlearn/fairlearn

Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0206-3

Knibbs, K. (2024, March 20). Here's proof you can train an AI model without slurping copyrighted content. *WIRED*. https://www.wired.com/story/proof-you-can-train-ai-without-slurping-copyrighted-content/

Norwest Venture Partners. (2024, January 24). *Ethical AI for startup Founders and Teams: Key questions and foundational steps - Norwest Venture Partners*. https://www.nvp.com/blog/ethical-ai-for-startup-founders/

PricewaterhouseCoopers. (n.d.). *Ten principles for ethical AI*. PwC. https://www.pwc.com.au/digitalpulse/ten-principles-ethical-ai.html

Sullivan, M. (2023, October 24). *Key principles for ethical AI development*. Transcend Blog. https://transcend.io/blog/ai-ethics

The Aletheia Framework®. (n.d.). https://www.rolls-royce.com/innovation/the-aletheia-framework.aspx