QUICKSCAN - CANVAS Synthetic Data Generation for Skin Cancer Detection

NAME: Synthetic Data Generation for Skin Cancer Detection

DATE: May 15, 2026 1:48 PM

DESCRIPTION OF TECHNOLOGY

This project builds a proof-of-concept AI pipeline that combines real dermatoscopic datasets with diffusion-based synthetic images to support skin-cancer detection. A Kedro backend and Streamlit UI enable data generation and model training to study privacy, fairness across skin types, and diagnostic performance.

HUMAN VALUES

Intended users are researchers and clinicians exploring synthetic data and semi-supervised learning, not patients. The tool can strengthen their sense of competence by making complex pipelines more approachable and experimental. At the same time it can encourage viewing patient images as technical resources, so I want that ethical tension to stay visible in documentation and discussions.

TRANSPARENCY

The idea, main workflows and data sources in the project documents and in the Streamlit POC interface. It is explicit that this is a personal research prototype, not a certified medical device, and that there is no commercial business model behind it. Known limitations, risks and open questions are described in documentations so stakeholders can interpret the results cautiously.

IMPACT ON SOCIETY

Medical AI for skin-cancer detection needs many high-quality labelled images, but collecting and annotating dermatology data is slow, expensive and tightly constrained by privacy regulations. This technology explores whether synthetic images and semi-supervised learning can ease that bottleneck and still reach clinically useful performance, so diagnostic AI becomes more accessible and cost-efficient.

STAKEHOLDERS

- Qin Zhao - Fontys Project Owner
- Ralf Raumanns - Dermatology / AI domain expert
- Study Research Group
- Dermatological AI researchers and students using the tool

SUSTAINABILITY

Training and running generative and classification models consumes GPU time and therefore energy. I try to limit this by re-using pre-trained models where possible, keeping experiment grids small and using shared research infrastructure instead of separate long-running servers. Future work could add simple tracking of compute per run to guide greener choices.

HATEFUL AND CRIMINAL ACTORS

Yes, under certain circumstances it could be misused. Synthetic or curated images and performance metrics could be falsely presented as if they came from fully validated clinical studies, helping bad actors market unapproved diagnostic tools or bypass proper regulatory and ethical review in other medical domains.

DATA

Yes. I assume datasets are incomplete, subjective and biased by clinic, device and selection effects. Synthetic images are always evaluated on real, held-out data, and I check for overfitting to artefacts or obvious domain shift. Limitations and remaining risks are documented together with the results, rather than treated as solved.

FUTURE

If approaches like this mature, synthetic dermatology data and semi-supervised learning could lower annotation costs and let more teams prototype diagnostic tools with less real patient data. At the same time, weak governance around synthetic data and validation could hide important performance gaps, leading to over-confident deployment of fragile models.

PRIVACY

The prototype uses existing dermatology image datasets that are already pseudonymised by their providers and then generates synthetic images from them. It does not collect new identifiers such as names or contact details. However, because the source images depict real lesions, I treat them conceptually as health-related personal data and align the design with GDPR-style principles.

INCLUSIVITY

Yes. Using popular benchmark datasets bakes in their class balance, clinical setting and demographic mix. Synthetic generation and model choices inevitably reflect those starting points and my own assumptions as a single researcher. Part of the project is to surface these biases through exploratory analysis and transparent reporting, not to pretend they are fully removed.

FIND US ON WWW.TICT.IO

THIS CANVAS IS PART OF THE TECHNOLOGY IMPACT CYCLE TOOL. THIS CANVAS IS THE RESULT OF A QUICKSCAN. YOU CAN FILL OUT THE FULL TICT ON WWW.TICT.IO

QUICKSCAN - CANVAS for Synthetic Data Generation for Skin Cancer Detection

NAME: Synthetic Data Generation for Skin Cancer Detection
DATE: May 15, 2026 1:48 PM

DESCRIPTION OF TECHNOLOGY
This project builds a proof-of-concept AI pipeline that combines real dermatoscopic datasets with diffusion-based synthetic images to support skin-cancer detection. A Kedro backend and Streamlit UI enable data generation and model training to study privacy, fairness across skin types, and diagnostic performance.

HUMAN VALUES

How is the identity of the (intended) users affected by the technology?

To help you answer this question think about sub questions like:

- If two friends use your product, how could it enhance or detract from their relationship?
- Does your product create new ways for people to interact?…

TRANSPARENCY

Is it explained to the users/stakeholders how the technology works and how the business model works?

- Is it easy for users to find out how the technology works?
- Can a user understand or find out why your technology behaves in a certain way?
- Are the goals explained?
- Is the idea of the technology explained?
- Is the technology company transparent about the way their…

IMPACT ON SOCIETY

What is exactly the problem? Is it really a problem? Are you sure?

Can you exactly define what the challenge is? What problem (what 'pain') does this technology want to solve? Can you make a clear definition of the problem? What 'pain' does this technology want to ease? Whose pain? Is it really a problem? For who? Will solving the problem make the world better? Are you sure? The problem definition will help you to determine…

STAKEHOLDERS

Who are the main users/targetgroups/stakeholders for this technology? Think about the intended context by…

When thinking about the stakeholders, the most obvious one are of course the intended users, so start there. Next, list the stakeholders that are directly affected. Listing the users and directly affected stakeholders also gives an impression of the intended context of the technology.
…

SUSTAINABILITY

In what way is the direct and indirect energy use of this technology taken into account?

One of the most prominent impacts on sustainability is energy efficiency. Consider what service you want this technology to provide and how this could be achieved with a minimal use of energy. Are improvements possible?

HATEFUL AND CRIMINAL ACTORS

In which way can the technology be used to break the law or avoid the consequences of breaking the law?

Can you imagine ways that the technology can or will be used to break the law? Think about invading someone's privacy. Spying. Hurting people. Harassment. Steal things. Fraud/ identity theft and so on. Or will people use the technology to avoid facing the consequences of breaking the law (using trackers to evade speed radars or using bitcoins to launder…

DATA

Are you familiar with the fundamental shortcomings and pitfalls of data and do you take this sufficiently into…

There are fundamental issues with data. For example:

- Data is always subjective;
- Data collections are never complete;
- Correlation and causation are tricky concepts;
- Data collections are often biased;…

FUTURE

What could possibly happen with this technology in the future?

Discuss this quickly and note your first thoughts here. Think about what happens when 100 million people use your product. How could communities, habits and norms change?

PRIVACY

Does the technology register personal data? If yes, what personal data?

If this technology registers personal data you have to be aware of privacy legislation and the concept of privacy. Think hard about this question. Remember: personal data can be interpreted in a broad way. Maybe this technology does not collect personal data, but can be used to assemble personal data. If the technology collects special personal data (like…

INCLUSIVITY

Does this technology have a built-in bias?

Do a brainstorm. Can you find a built-in bias in this technology? Maybe because of the way the data was collected, either by personal bias, historical bias, political bias or a lack of diversity in the people responsible for the design of the technology? How do you know this is not the case? Be critical. Be aware of your own biases.…

FIND US ON WWW.TICT.IO

THIS CANVAS IS PART OF THE TECHNOLOGY IMPACT CYCLE TOOL. THIS CANVAS IS THE RESULT OF A QUICKSCAN. YOU CAN FILL OUT THE FULL TICT ON WWW.TICT.IO

Fontys University of Applied Sciences

TICT