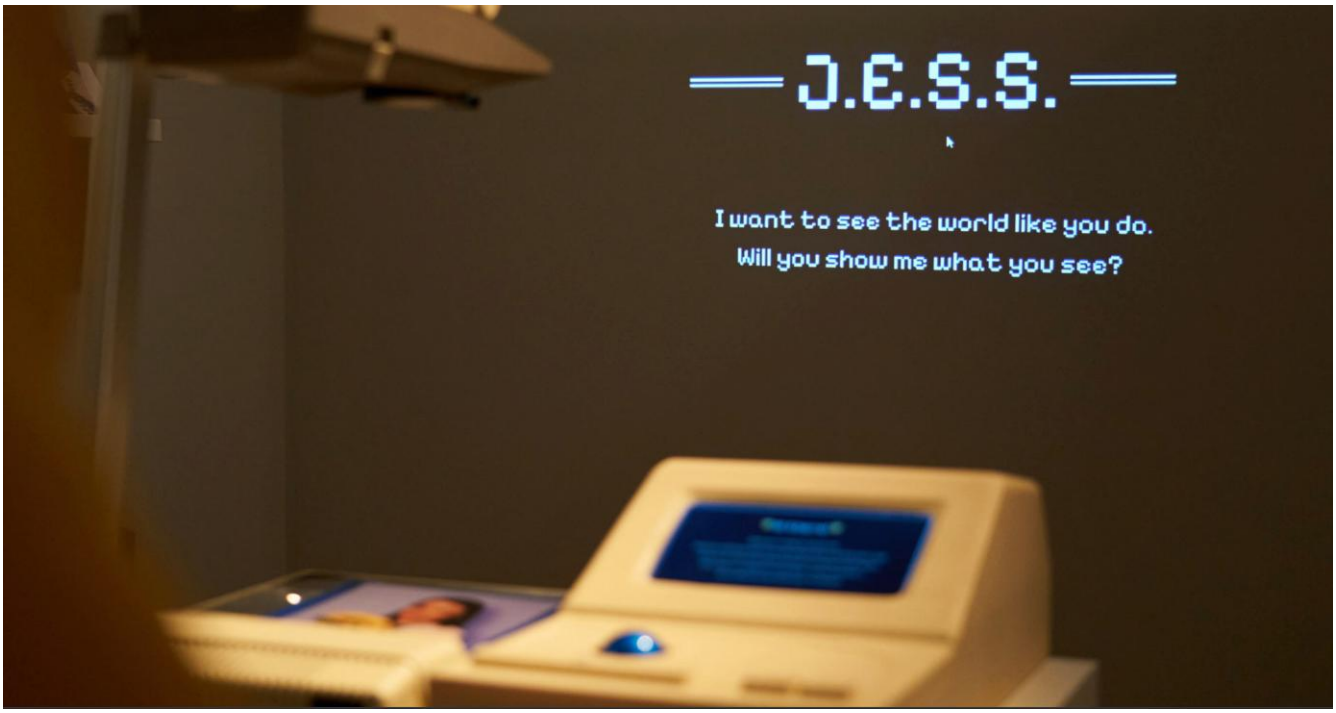



# J.E.S.S.

*Just Examining Systemic Stereotypes*

## Final Project Report

Submitted to Bedford Creative Arts

Artists: Arnab Chakravarty & Fergus Laidlaw
Report date: February 2025

# 1. Project Overview

J.E.S.S. (Just Examining Systemic Stereotypes) is an interactive art installation commissioned by Bedford Creative Arts as part of the Community Portrait exhibition at The Higgins Bedford. The work was created by artists Arnab Chakravarty and Fergus Laidlaw, with production support from Ami Aubrey and Bex Foskett.

The central provocation of the work is articulated in the following statement, which anchored the exhibition label and all promotional materials:

> *"Portrait photography creates images by elevating humans beyond their stereotypes, while AI image models generate images by flattening humans to their stereotypes."*

J.E.S.S. makes visible the mechanics and biases of AI image generation by taking visitors through the full pipeline: a human-taken portrait photograph is fed into a hybrid projector-scanner machine, translated by an AI image-to-text model (CLIP) into language, and then reconstructed into a new image by an AI text-to-image model (Stable Diffusion, via ComfyUI). The resulting outputs are displayed side-by-side with the original photographs, with the intermediate machine-generated text visible for visitors to read.

The installation was deliberately housed in the casing of a retrofitted overhead projector — a nostalgic, tactile object — to make the often-abstract concept of AI approachable and legible to a general public audience. Visitors were invited to select photographs themselves, initiate the process, and reflect on what the machine noticed, missed, or distorted. Pen-and-paper response sheets were placed alongside the machine, inviting visitors to record their own words in response to what they observed.

The Community Portrait exhibition provided the ideal context for this work: over 200 portraits of Bedford residents, created through artist-led workshops in schools, community settings, Bedford College, and The Higgins Bedford, offered a diverse and authentic repository of human images to test against the AI system's assumptions.

Over the course of the exhibition, these visitor-written responses became one of the most striking outputs of the project. Where the AI system produced labels, visitors produced meaning — writing about emotion ("love", "shy confidence", "staring into the future"), relationships ("a father and his children", "holds them both tight", "a couple in love"), and correcting the machine directly when it got things wrong ("showed a white man instead", "dressed differently").

These responses point to something fundamentally reassuring: the biases J.E.S.S. exposes in AI image generation are not biases that humans share. Given the opportunity, people instinctively reached for the specific, the relational, and the felt. The machine could label a face; visitors described a person. The sections that follow document the exhibition's technical delivery, public engagement, and a detailed comparative analysis of these two bodies of language.

# 2. Exhibition & Presentation History

## 2.1 Community Portrait Exhibition — The Higgins Bedford

Venue: Sir William Harpur Gallery, The Higgins Bedford, Castle Lane, Bedford MK40 3XD

Dates: 28 September 2024 – 23 February 2025

Entry: Free

J.E.S.S. was installed as a featured interactive element within the Community Portrait exhibition, positioned in dialogue with the broader arc of photographic history presented in the gallery — from 1840s glass-plate techniques through to contemporary digital photography. The installation functioned as the exhibition's speculative forward gesture: a provocation about where image-making might go, and what is at stake when AI tools are used to represent human beings.

## 2.2 BEYOND 2024 Conference — Salford/Manchester

Venue: BEYOND 2024 Conference, Manchester

Dates: 25–27 November 2024

J.E.S.S. was presented at the BEYOND 2024 conference, a national gathering focused on digital arts, participation, and technology. The artists attended alongside the Bedford Creative Arts team, bringing the installation directly from The Higgins Bedford for the duration of the event. Presented to an audience of arts professionals, digital practitioners, and cultural producers, the installation generated significant interest and was mentioned in the closing keynote of the conference.

## 2.3 BCA 40th Anniversary Celebration

Venue: BizSpace, Bedford Heights, Bedford

Date: 13 November 2025

J.E.S.S. was selected as one of the featured demonstrations at Bedford Creative Arts' 40th anniversary celebration, attended by approximately 100 guests including artists, community leaders, trustees, funders, and past and present BCA collaborators. Arnab Chakravarty presented a live demonstration of the installation alongside Dr Atif Ghani's demonstration of the Airships XR project, making J.E.S.S. one of two digital artworks chosen to represent BCA's current programme.

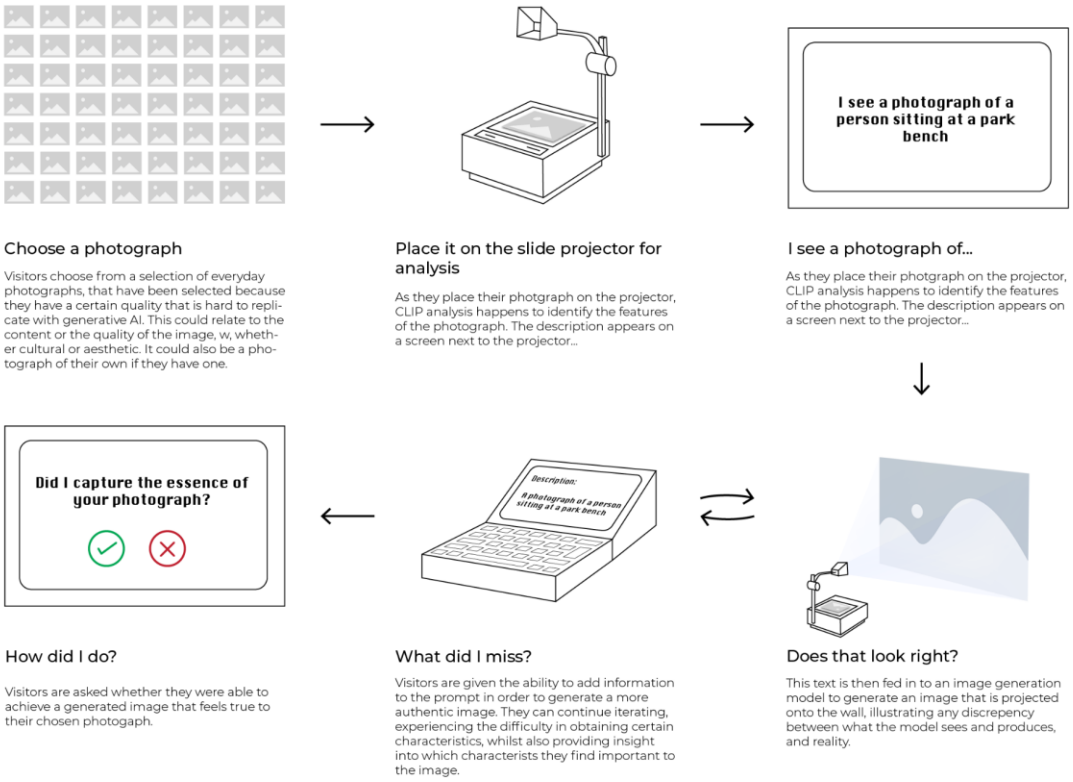
## 2.4 Reprint Revolution Symposium

Venue: Bedford Creative Arts

Date: 14 January 2025

On 14 January 2025, Bedford Creative Arts hosted Community Portrait: Reprint Revolution, a one-day public symposium examining the intersection of AI and community image-making. Arnab Chakravarty and Fergus Laidlaw spoke as part of a panel programme also featuring Bedford College lecturer Grete Dalum-Tilds, digital participation expert John Whall, and artist-educator Andee Collard. The symposium offered a more formal critical framing for the questions J.E.S.S. raises, and drew together practitioners and community members to discuss algorithmic bias, representation, and the creative potential of AI in socially engaged art.

# 3. Technical Delivery




## 3.1 System Architecture

J.E.S.S. runs on a custom ComfyUI pipeline housed within a retrofitted overhead projector. When a visitor places a photograph on the projector bed, a camera embedded in the projector arm captures the image and passes it to a CLIP image-to-text model, which generates a set of descriptive tags. These tags are displayed on the wall projection alongside the original photograph under the heading "I see..." — making the machine's interpretation immediately legible and comparable to the human image it is describing. The tags are then passed to Stable Diffusion, which generates a new image from them, also projected onto the wall. Visitors can indicate which tags they felt were accurate or inaccurate, and are invited to write their own words on a physical card placed alongside the machine. The iterative loop — photograph, machine description, generated image, visitor response — is the core interactive and analytical structure of the work.

The physical enclosure was built around a retrofitted overhead projector chassis, presenting the system as a single self-contained unit. The installation is designed to be plug-and-play, requiring only a 2m x 2m wall surface for projection, a standard 230V power supply, and positioning the projector unit 2–3m from the wall.


## 3.2 Technical Specifications

Software: P5.js, ComfyUI

AI models: CLIP (image-to-text), Stable Diffusion (text-to-image)

User interaction: Photograph placed on projector bed; camera in projector arm captures image; visitors respond via tag selection and handwritten cards

Generation time: Approximately 90 seconds per image

### 3.3 Technical Observations

J.E.S.S. ran entirely on local hardware with no data sent to the cloud at any point during its operation, an important consideration both technically and ethically given the nature of the community photographs being processed. As an experimental setup running multiple AI models on cost-effective consumer hardware, the installation encountered some technical instability during its initial days at The Higgins Bedford. Through iterative adjustment over the course of the exhibition, the system stabilised and ran reliably without interruption for the remainder of its run.

The central technical achievement of J.E.S.S. is the compression of a full AI pipeline onto affordable, self-contained hardware, covering image capture, CLIP inference, and Stable Diffusion generation. This makes the work genuinely portable and economically accessible to host, which is directly relevant to its potential as a touring installation.

## 4. Public Engagement

### 4.1 Exhibition Reach

The Community Portrait exhibition, of which J.E.S.S. was a central component, welcomed over 5,000 visitors during its five-month run at The Higgins Bedford. As a free, publicly accessible exhibition, it reached a broad cross-section of the Bedford community, including many of the individuals and community groups whose portraits were included in the show.

### 4.2 Visitor Interaction (The Higgins Bedford):

Most visitors arrived with little awareness of AI image generation, which meant the installation required a moment of orientation before its full effect landed. Once the process clicked, the response was consistently one of awe — visitors were struck both by how close the generated image got to the original photograph and by how plainly it exposed the system's tendency to stereotype. One unintended interaction emerged repeatedly: visitors wanted to try the machine on their own photographs, holding their phones beneath the projector arm to feed in personal images. The emotional response when seeing themselves processed through the system was considerably more intense than with the curated portrait set.

### 4.3 Conference and Specialist Audience (BEYOND 2024):

At BEYOND, a technically literate audience engaged with J.E.S.S. quite differently — already familiar with generative AI, visitors experimented more strategically, deliberately selecting photographs they anticipated would surface stereotypical outputs and testing the system across multiple images. The same impulse to use personal photographs emerged here too, and with similar intensity.

Enabling this interaction responsibly remains an open design challenge: as a probabilistic system, J.E.S.S. cannot guarantee its outputs, raising real questions about how to allow personal image input without risking offensive or unintended results.

## 5. Data & Findings: Machine Language vs. Human Language

J.E.S.S. generated two parallel bodies of text over the course of its exhibition: the machine-produced CLIP tags describing each photograph, and the handwritten words left by visitors on response cards. The contrast between these two corpora is the evidential core of the work's argument.

## 5.1 Machine Language: Patterns and Biases

The machine's vocabulary falls into four distinct patterns, each revealing a different dimension of bias. The largest category is ethnic and cultural labelling, where CLIP consistently reached for ethnicity as a primary descriptor when encountering non-western subjects. Terms such as "slavic women", "south asian", "muslim queen", "light blue saree", "bindi", "british police" and "west slav" demonstrate how the model defaults to cultural classification rather than individual description. Notably, not many ethnic labels appeared for subjects read as white and western. The model focused instead on physical attributes such as "middle aged woman."

The second most striking pattern is the model's reliance on commercial and stock photography references. Tags such as "500px", "pexels", "ko-fi", "instagram influencer", "8k portrait photography" and "catalog image" reveal that CLIP has been trained predominantly on commercial image platforms, interpreting human photographs through that aesthetic framework.

Celebrity anchoring formed a third category, where the model mapped subjects onto named figures and artists including "Billie Eilish aesthetic", "mia khalifa", "Annie Leibovitz", "Wes Anderson aesthetic" and "bollywood adjacent", flattening individual identity into cultural shorthand.

Finally, even the model's more neutral descriptors such as "surface blemishes", "sensual cold", "unconventional beauty" and "dark femme" carried the aesthetic register of commercial imagery, framing ordinary human features through the language of editorial fashion photography.

## 5.2 Human Language: What Visitors Noticed

The visitor-written responses are organised across five categories, and their character is fundamentally different from the machine's output. Where the machine produced labels, visitors produced language that was relational, emotional, narrative, and corrective.

The largest categories are emotional and observation. Visitors wrote "love", "longing", "shy confidence", "staring into the future", "quiet moment in time", "curiosity in her eyes", "unwatched, observing." These are not descriptors in any taxonomic sense but interpretations of humans that the machine had no equivalent vocabulary for.

The relational category is perhaps the starkest contrast: "a father and his children", "mother and daughter", "friends or sisters", "holds them both tight", "turned inward toward one another", "a couple in love." CLIP produced no relational language across its entire dataset. It could not identify that two people in a photograph might know each other.

The personal and political category showed how visitors brought their own contexts into the encounter: "Ukraine", "ordinary people like my family", "how much it has cost", "we should welcome them", "support to her motherland." These responses locate individual photographs within broader social and political frames, a register entirely absent from the machine's output.

The correction category recorded direct moments of misidentification: "showed a white man instead", "thought it was like me", "facing away when it was not", "it saw a large moustache", "dressed differently." These notes document specific instances where visitors identified the model's assumptions about race, gender, and appearance.

## 5.3 Comparative Analysis

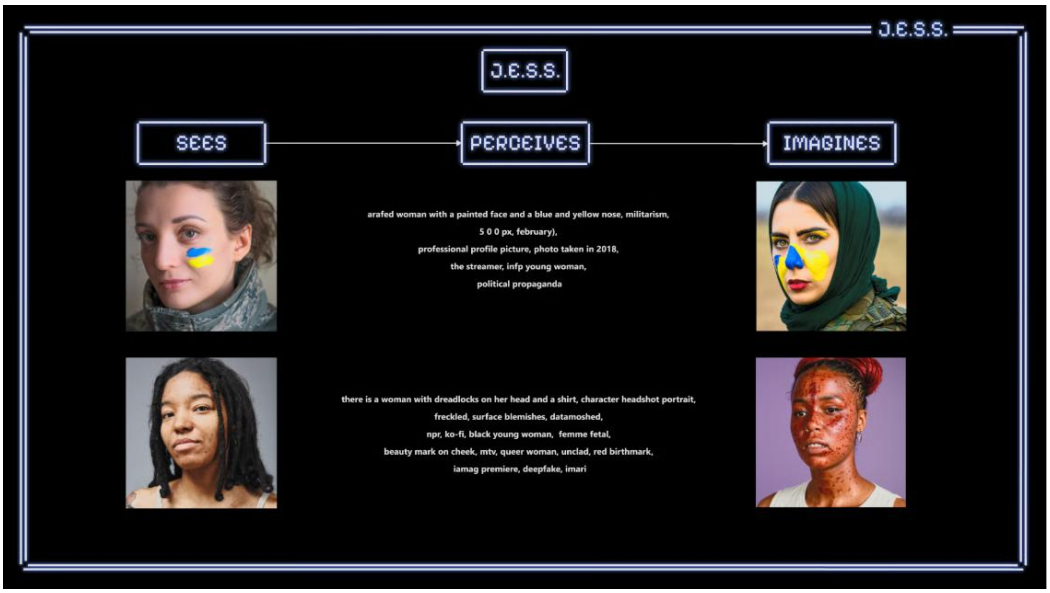
The gap between machine and human language is not merely a difference in vocabulary but a difference in what each considers worth saying about a human being. The machine's language is external, categorical, and commercially anchored. The visitor's language is internal, relational, and contextual.

What J.E.S.S. makes visible is that these biases are structural features of how the model was built. CLIP was trained on commercially curated datasets and sees the world through that lens: stripped of relationship, flattened into type, and optimised for searchability. A Black woman with dreadlocks produces tags including "surface blemishes", "datamoshed", "npc", "ko-fi", "femme fatal" and "deepfake" — a set that combines clinical notation, internet subculture references, and implicitly loaded framing, none of which a human visitor applied to the same image.

The correction cards are the most precise evidence in the dataset. "Showed a white man instead" is not only a note about a technical failure but a record of the model substituting one person for another based on its training assumptions. These moments of misidentification were not isolated. They followed consistent patterns around race, gender, and cultural markers.

The intermediate text layer — the CLIP tags displayed between the original photograph and the generated image — proved to be the most analytically significant element of the work. Language made the model's assumptions legible in a way that the generated image alone did not.

## 5.4 Exemplar Analysis




The curated photographs above illustrate the patterns identified in the preceding sections with particular clarity and are worth examining together as a cohort.

**Substitution:** Across several examples, the model replaces individual context with cultural stereotypes. A woman in a hijab and hi-vis vest holding a hard hat is anchored to "mia khalifa", "mo salah" and "moorish architecture". The professional context disappears and the hijab becomes the dominant interpretive signal. A woman with Ukrainian flag face paint is described as "militarism" and "political propaganda", her act of solidarity reframed as a threat. A Black woman with dreadlocks in a plain shirt produces "femme fetal", "unclad", "deepfake" and "datamoshed", terms that introduce connotations entirely absent from the original image. In each case the generated output reflects the stereotype encoded in the intermediate text rather than the photograph that was scanned.

**Erasure:** Two examples show the model encountering something outside its trained categories and resolving it away rather than representing it. A young woman with a prosthetic arm taking a selfie produced tags including "glass arms" and "wrapped arms". The generated image removed the prosthetic entirely, rendering her with two standard arms. A professional headshot of a smiling woman in a blue jacket produced tags including "transgender", "uncanny valley" and "similar to hagrid". The model appears to have associated certain facial features or presentation with a gender identity, applying it as a descriptor with no visible basis in the image. The generated output then resolved this back into a more conventionally feminine figure, older and more heavily styled. Where the intermediate text introduced a reading outside a binary gender category, the image generation stage corrected for it. The system has no mechanism for representing gender outside its trained categories.

**Corruption:** A recurring technical artefact across multiple examples is the prefix "arafed", appearing as "arafed woman in hijab" and "arafed woman with a painted face." This appears to be a corrupted token in CLIP's training data that the model consistently applies to certain subjects. Its presence across unrelated photographs points to a systematic error embedded in the model's vocabulary that compounds the interpretive biases described above.

**Synthesis:** Taken together, these examples demonstrate that the model's errors are not random. They follow consistent structural patterns: cultural markers override individual context, differences that fall outside trained categories are detected and then erased, and corrupted training data

introduces additional distortions. The intermediate text layer makes each of these moves legible in a way that the generated image alone would not.

### 5.5 Data Visualization

The accompanying data visualisation presents the machine and human word datasets as a live animated field, divided vertically down the centre of the screen. Machine-generated tags occupy the left half, visitor-written words the right. The split format offers an accessible way to understand at a glance how the machine and human responses to the same photographs diverge in character, register, and content. Patterns that emerge from reading the two sides together reinforce the findings of the preceding analysis: the machine produces labels, the human produces meaning.


The visualisation can be viewed at: https://editor.p5js.org/AbolTaabol/full/2CD9RShnX

## 6. Artistic & Critical Reflections

The development and presentation of J.E.S.S. over the course of the Community Portrait project has generated a range of artistic and critical insights — both in terms of what the work revealed about AI image generation, and what it revealed about how audiences engage with these systems when given direct, embodied access to them.

### 6.1 What the Work Revealed

Beyond the patterns of bias documented in the preceding sections, the exhibition surfaced several other findings worth noting:

The system proved brittle in the presence of ambient lighting, shadows, glares and reflections. When the model encountered ambiguity in the image, it consistently resolved it toward the stereotypical, as if uncertainty pushed the probabilistic output further into its trained defaults rather than toward openness.

The visitor-written responses were, unexpectedly, among the most striking material the work produced. Written quickly and on instinct, they were often quietly poetic and precise, in sharp contrast to the machine's output. Where algorithmic systems and digital platforms routinely sort

people into categories, visitors instinctively chose not to. That impulse toward the specific, the relational, the felt appeared to be a default of human perception that the machine entirely lacked.

The slowness of the generation process, a product of the technical constraints described in Section 3, proved unexpectedly valuable as a design condition. In a landscape dominated by fast, frictionless AI generation, the enforced pause created space for contemplation. For audiences not already immersed in these technologies, slowness may offer a more productive way to engage with and understand generative systems than speed does.

Finally, the rapid pace of development in generative AI means that many of the specific failure modes J.E.S.S. surfaces are being obscured by newer, more capable models. The artists believe this makes the underlying analysis more rather than less important. Using fundamental models for examination offers a way to understand how these systems are innately trained, before newer capabilities paper over the evidence. The work asks audiences not to accept these systems as magic, to question their outputs, and to demand evaluations that consider whether the most vulnerable people in society are being unfairly harmed.

## 6.2 Audience and Context

The two primary contexts drew out different but complementary dimensions of the work. At The Higgins Bedford, the five-month duration allowed a gradual accumulation of visitor responses, and the general public audience engaged with the work without prior agenda. Visitors arrived without strategic intent, chose photographs that caught their eye, and responded with genuine surprise.

At BEYOND, the specialist audience tested the installation's limits deliberately, selecting photographs they expected would expose stereotyping and iterating across multiple images. This produced a more analytical engagement but confirmed the same underlying findings. The biases it exposed were real and consistent, not artefacts of particular photographs or isolated failures. Analysing the collected data after the exhibition confirmed the original hypothesis, though the clarity and consistency with which the bias manifested went beyond what was anticipated.

One unintended finding across both contexts was the impulse visitors had to feed their own photographs into the machine. The emotional response was considerably more intense than with the curated set, suggesting the work's argument landed differently when it became personal. Visitors using their own images often responded with humour, treating the machine's misreadings as playful provocation, pointing to potential applications in more informal social settings where the same biases surface through play rather than critical engagement.

## 6.3 The Role of the Physical Object

The decision to house J.E.S.S. in the casing of a retrofitted overhead projector proved central to the work's reception. The familiar, nostalgic object lowered the threshold for engagement, making the AI system feel approachable rather than intimidating or opaque. Visitors who might have been deterred by a screen-based interface were drawn in by the physicality of placing a photograph on the projector bed — a gesture that felt intuitive, deliberate, and considered in a way that a touchscreen interaction would not have.

The overhead projector also carried its own cultural associations: education, shared looking, the authority of the projected image. Housing a generative AI system inside that object created a productive tension between what the machine appeared to be and what it actually did. The retro exterior and the generative AI system inside share more than a housing. Both look to the past to

project into the future — the overhead projector through the literal throw of light onto a wall, generative AI through the patterns encoded in its training data. This parallel foregrounded one of the work's central questions: what technology remembers, and what it forgets.

## 7. Key Links

### Press:

https://www.digitalcameraworld.com/news/fujifilm-supported-exhibition-provides-futuristic-glimpse-into-ai-wild-west

https://www.bbc.co.uk/news/articles/cpqz913z18qo

### Project details:

https://bedfordcreativearts.org.uk/events/community-portrait-reprint-revolution/

https://bedfordcreativearts.org.uk/articles/bca-at-beyond/

https://bedfordcreativearts.org.uk/articles/celebrating-40-years-of-bca/

https://beyondconference.org/archive/b24/showcase/bedford-creative-arts-ai-community-portrait.html

https://www.ferguslaidlaw.com/work/community-portraits

https://chaky.space/databases/work-database/jess

### Visualizer:

https://editor.p5js.org/AbolTaabol/full/2CD9RShnX

## 8. Acknowledgements

J.E.S.S. was made possible through the support and commitment of a number of partners and individuals.

Commissioner: Bedford Creative Arts

Partner venue: The Higgins Bedford

Principal sponsor: Fujifilm UK

Funders: Arts Council England, The Harpur Trust, The Wixamtree Trust, The Steel Charitable Trust, Bedford Borough Council

The artists would also like to thank the following individuals for their contributions to J.E.S.S.:

Producer: Ami Aubrey

Logistics & Installation: Bex Foskett

Operations: Lydia Saul & Higgins team

Finally, the artists extend their sincere thanks to all the members of the Bedford community whose portraits animated this work, and to every visitor who took the time to place a photograph into the machine and record what they saw.

# 9. Copyright and Enquiries



J.E.S.S. (Just Examining Systemic Stereotypes) was commissioned by Bedford Creative Arts. For all enquiries relating to J.E.S.S. — including exhibition loans, touring, licensing, press, or collaboration — please contact Bedford Creative Arts: bedfordcreativearts.org.uk