

INTERACTIVE & EXPLAINABLE AI

Nikola Banovic

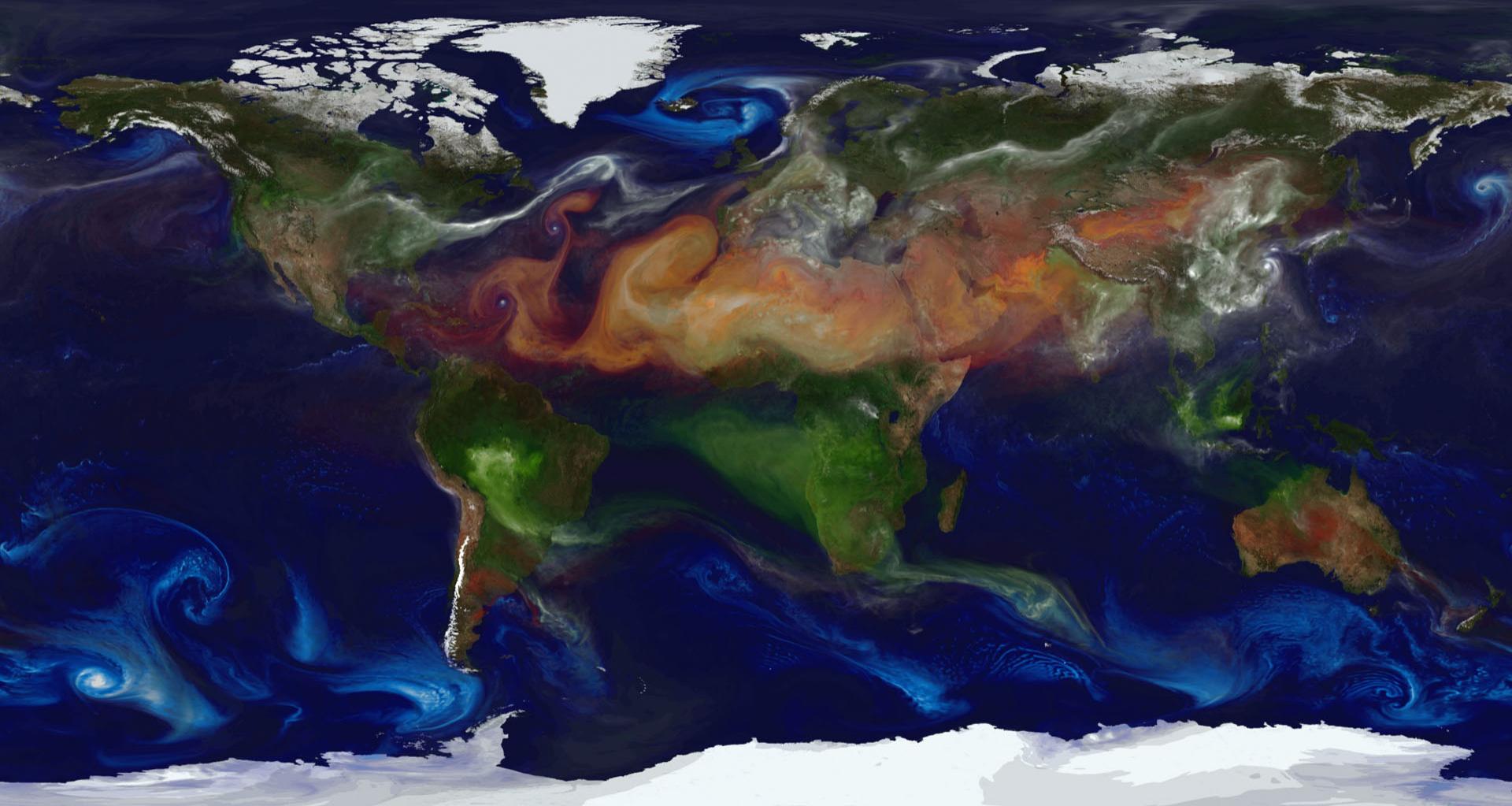
Assistant Professor
Computer Science & Engineering
University of Michigan
nbanovic@umich.edu
<http://www.nikolabanovic.net>

The 6th Summer School on Computational Interaction – Day 4

ABOUT ME

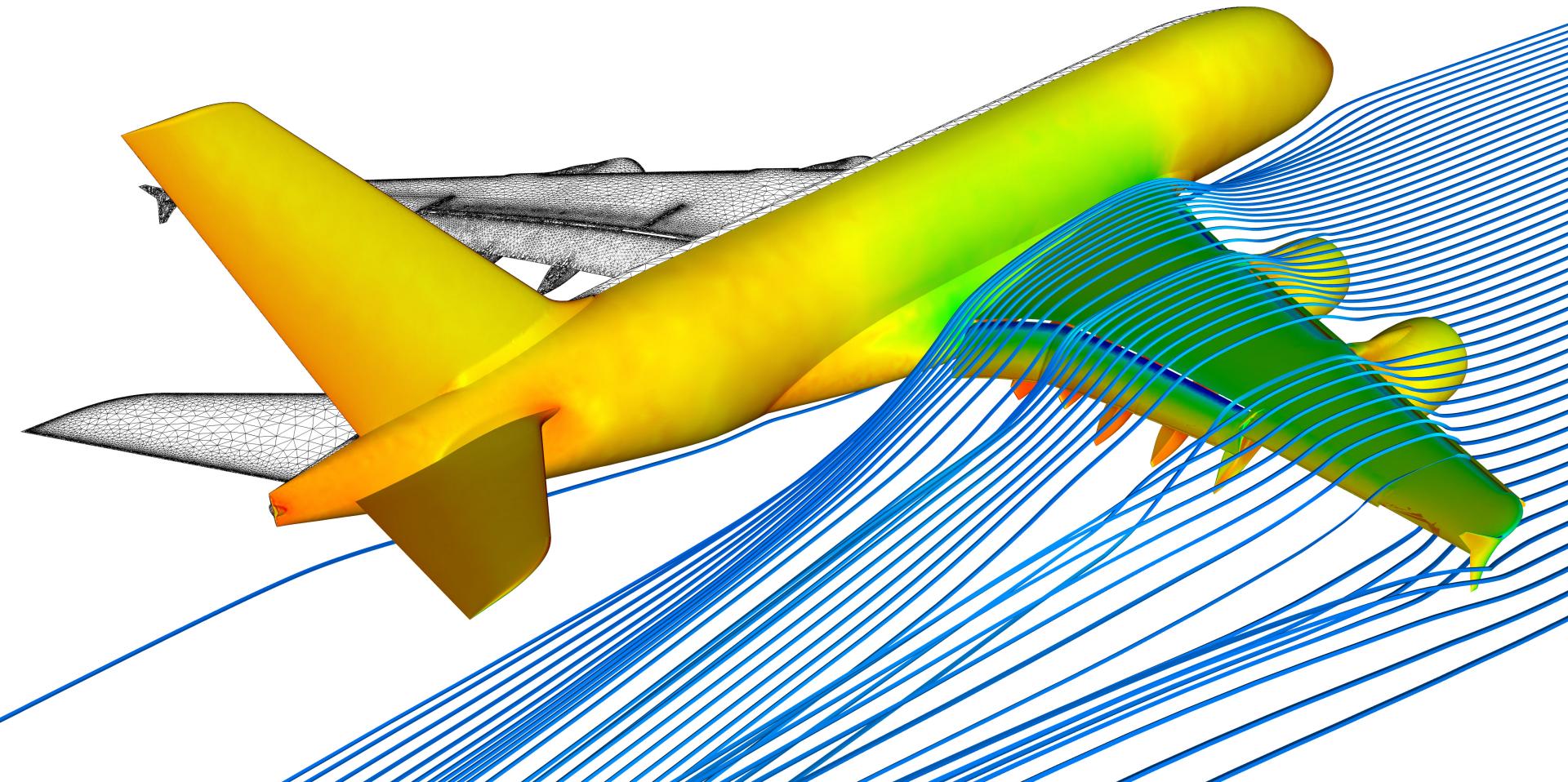
- Assistant Professor at the Computer Science and Engineering department at the University of Michigan
- PhD in Human-Computer Interaction from CMU
- Computer scientist doing HCI research; focus on Computational Interaction
- Research Interests:
 - Human Behavior Modeling
 - Explainable and Interpretable AI
 - Behavior-aware User Interfaces

COMPUTATIONAL MODELING: A TOOL TO STUDY AND EXPLORE COMPLEX SYSTEMS



View of aerosol movement created by NASA's models and supercomputers. NASA/Goddard Space Flight Center.

COMPUTATIONAL MODELING: A TOOL TO STUDY AND EXPLORE COMPLEX SYSTEMS



HUMAN-COMPUTER INTERACTION STUDIES PEOPLE'S INTERACTION WITH TECHNOLOGY



MODELING POINTING AND TYPING BEHAVIORS

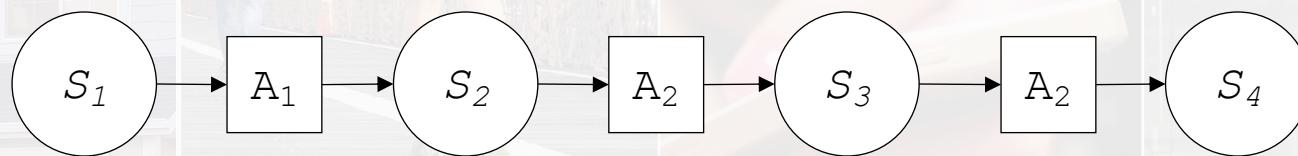


Banovic, Sethapakdi, Hari, Dey, Mankoff. 2019. The Limits of Expert Text Entry Speed on Mobile Keyboards with Autocorrect. In *Proc. MobileCHI '19*. ACM.

 Banovic, Rao, Saravanan, Dey, Mankoff. 2017. Quantifying Aversion to Costly Typing Errors in Expert Mobile Text Entry. In *Proc. CHI '17*. ACM. **Honorable Mention Award**

Banovic, Grossman, & Fitzmaurice. 2013. The Effect of Time-based Cost of Error in Target-directed Pointing Tasks. In *Proc. CHI '13*. ACM.

MODELING COMPLEX ROUTINE BEHAVIORS

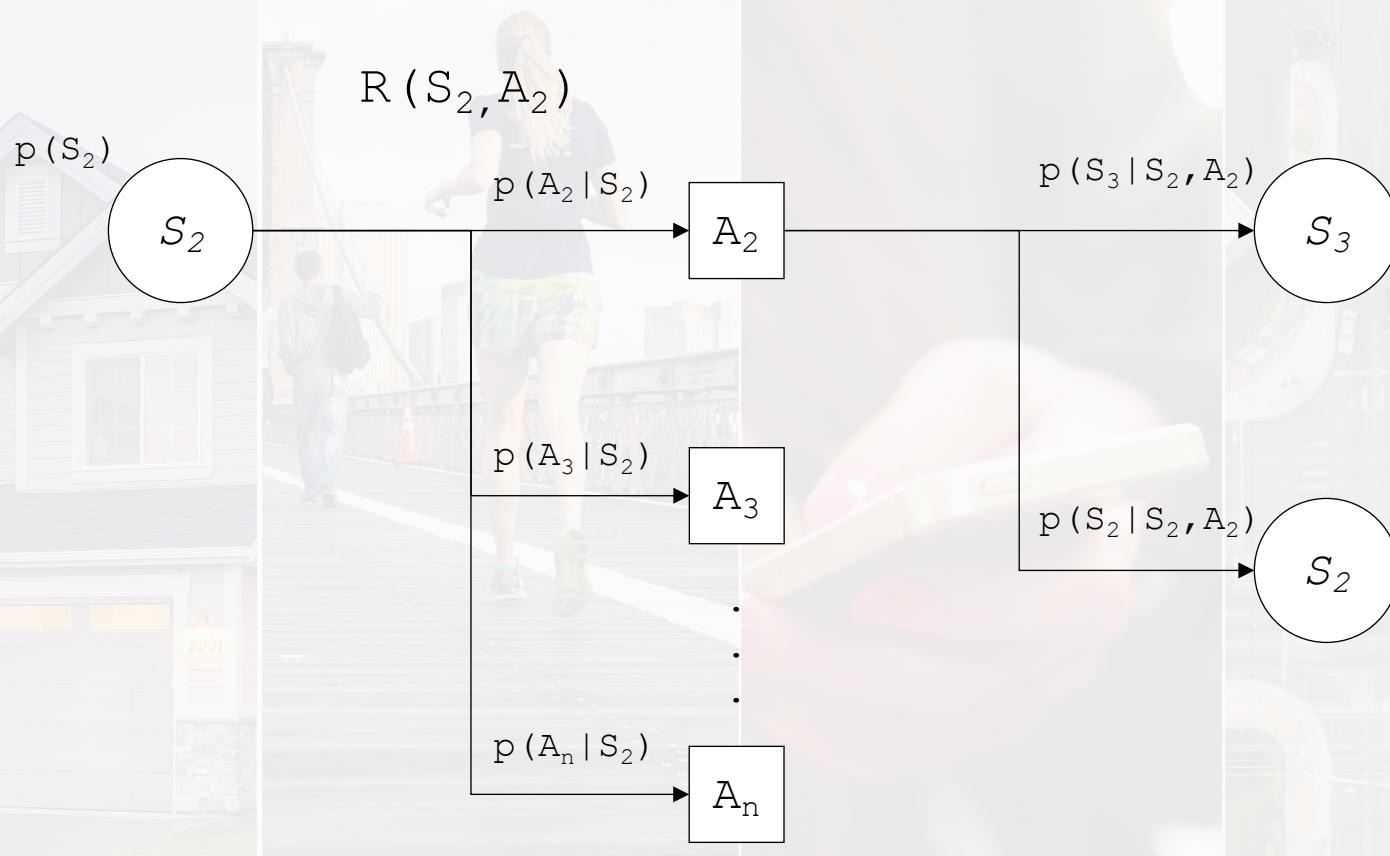


Banovic, Mankoff, and Dey. 2018. Computational Model of Human Routine Behavior. In *Computational Interaction*. Oxford University Press.

Banovic, Wang, Jin, Cheng, Ramos, Dey, & Mankoff. 2017. Leveraging Human Routine Models to Detect and Generate Human Behaviors. In *Proc. CHI '17*.

 Banovic, Buzali, Chevalier, Mankoff, & Dey. 2016. Modeling and Understanding Human Routine Behavior. In *Proc. CHI '16*. **Honorable Mention Award**

MODELING COMPLEX ROUTINE BEHAVIORS



Banovic, Mankoff, and Dey. 2018. Computational Model of Human Routine Behavior. In *Computational Interaction*. Oxford University Press.

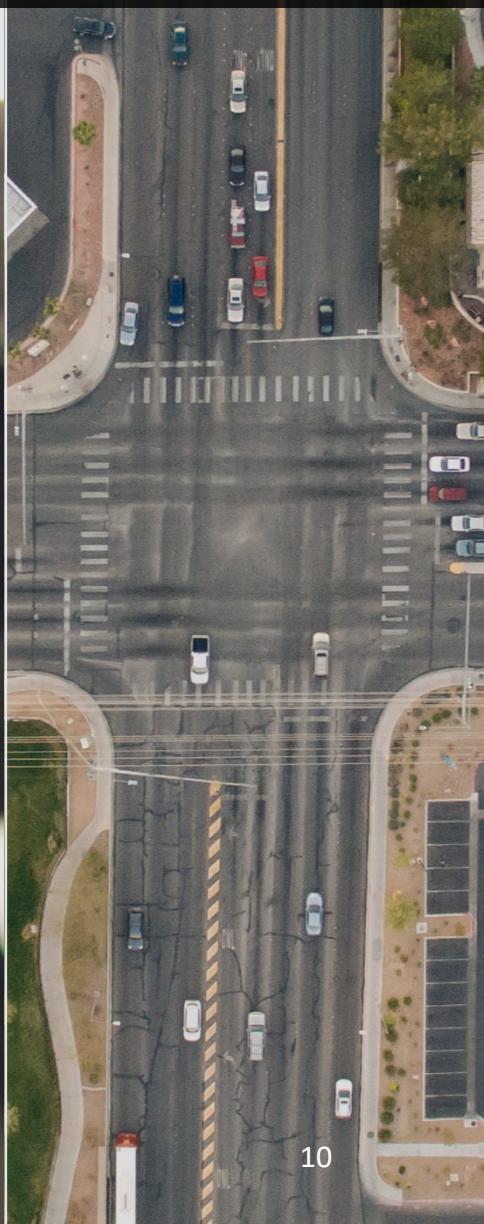
Banovic, Wang, Jin, Cheng, Ramos, Dey, & Mankoff. 2017. Leveraging Human Routine Models to Detect and Generate Human Behaviors. In *Proc. CHI '17*.

 Banovic, Buzali, Chevalier, Mankoff, & Dey. 2016. Modeling and Understanding Human Routine Behavior. In *Proc. CHI '16*. **Honorable Mention Award**

SENSE, COLLECT, AND STORE DATA ABOUT PEOPLE AT SCALE



NEW SOURCE OF DATA TO STUDY BEHAVIORS ABOUT MANY ASPECTS OF PEOPLE'S LIVES



DRIVER COACH DETECTS AGGRESSIVE DRIVING & SIMULATES NON-AGGRESSIVE ALTERNATIVES



DRIVER COACH DETECTS AGGRESSIVE DRIVING



WHY EXPLAINABILITY MATTERS TO ME

- To use such computational models to understand people, the models we use need to be **explainable**
- To deploy behavior-aware user interfaces we need to ensure that we can **assess the capabilities and limitations** of such systems

LEARNING GOALS

- Define Explainability and Interpretability
- Learn what we mean by Explainable AI (XAI) including Human-Centered XAI (HCXAI)
- Learn about interactive model exploration (a particular HCXAI method)
- Take a deep dive into an example interactive model exploration implementation

DAY 4 SCHEDULE

9:30 – 11:00

Lecture and discussions
(will have a coffee break)

11:00 – 11:30

Go over GAN Explorer code
(see the github repository)

13:30 – 15:00

Mini hackathon

15:30– 16:30

Showcase

WHAT IS ARTIFICIAL INTELLIGENCE (AI)?

Even code: #2426146

<https://app.sli.do/event/ut4dDWikQr64YLzNqBDHVr>

You have 120 seconds...

DONE!

WHAT IS ARTIFICIAL INTELIGENCE (AI)?

“Intelligence demonstrated by machines.”

WHAT IS EXPLAINABLE AI?

Even code: #2426146

<https://app.sli.do/event/ut4dDWikQr64YLzNqBDHVr>

You have 120 seconds...

DONE!

WHAT IS EXPLAINABLE AI?

“AI that humans can learn how it works, and know its capabilities and limitations.”

“AI that can provide reasons or justifications for its actions, decisions, or beliefs.”

WHAT IS INTERPRETABLE AI?

Even code: #2426146

<https://app.sli.do/event/ut4dDWikQr64YLzNqBDHVr>

You have 120 seconds...

DONE!

WHAT IS INTERPRETABLE AI?

“AI with inputs and outputs that can be translated into something that humans can understand.”

EXPLAINABLE ARTIFIAL INTELIGENCE (AI)

Explainability is of critical importance to technology-driven innovation and broader societal adoption of AI.

EXPLAINABLE ARTIFIAL INTELIGENCE (AI)

- The ability to explain capabilities and limitations of an AI and justify its decisions aids in its trustworthiness, accountability, fairness, inclusivity, and accessibility.

EXPLAINABLE AI (XAI)

- Existing approaches assume the user is a math-savvy model designer rather than a domain expert or an end-user without Computer Science training

Carvalho et al. 2019. Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics* 8, no. 8: 832. <https://doi.org/10.3390/electronics8080832>

Gilpin et al, 2018. Explaining Explanations: An Overview of Interpretability of Machine Learning," DSAA 2018, pp. 80-89. doi: 10.1109/DSAA.2018.00018.

Ras et al. 2018. Explanation Methods in Deep Learning: Users, Values, Concerns and Challenges. In Explainable and Interpretable Models in Computer Vision and Machine Learning. https://doi.org/10.1007/978-3-319-98131-4_2

EXPLAINABLE AI (XAI)

- Existing approaches assume the user is a math-savvy model designer rather than domain experts or end-users without Computer Science training
- Such explanations often do not match the end-users' mental models; thus, are ineffective and at times even harmful

Danding et al. 2019. Designing Theory-Driven User-Centric Explainable AI. In CHI '19, Paper 601, 1–15.
<https://doi.org/10.1145/3290605.3300831>

Alqaraawi et al, 2020. Evaluating saliency map explanations for convolutional neural networks: a user study. In IUI '20, 275–285. <https://doi.org/10.1145/3377325.3377519>

Lakkaraju et al. 2020. "How do I fool you?": Manipulating User Trust via Misleading Black Box Explanations. In Proc. of the AAAI/ACM Conference on AI, Ethics, and Society, 79–85. <https://doi.org/10.1145/3375627.3375833>

EXPLAINABLE AI (XAI)

- Existing approaches assume the user is a math-savvy model designer rather than domain experts or end-users without Computer Science training
- Such explanations often do not match the end-users' mental models; thus, are ineffective and at times even harmful
- Instead, enable end-users to explore capabilities and limitations of AI (and AI-based systems) through interaction that is tailored for them

WHAT ARE SOME EXAMPLES OF XAI?

Even code: #2426146

<https://app.sli.do/event/ut4dDWikQr64YLzNqBDHVr>

You have 120 seconds...

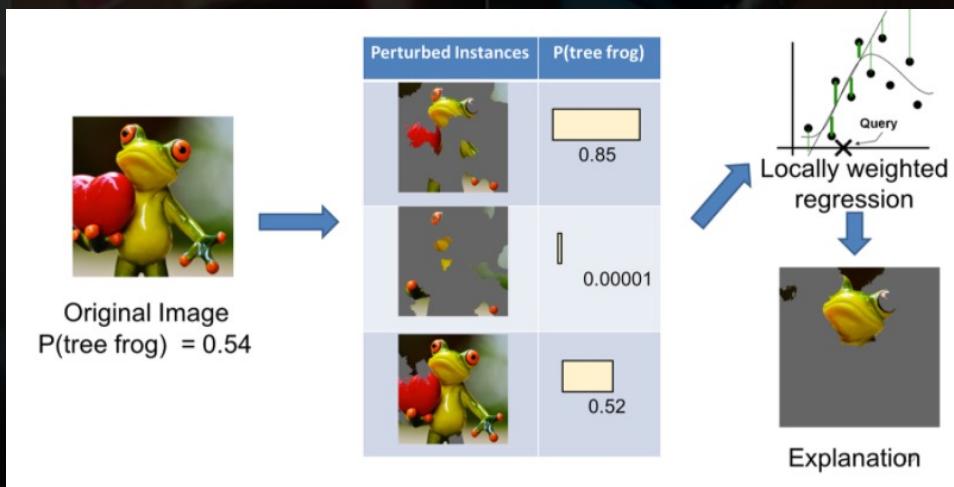
DONE!

EXAMPLE XAI: LIME

- Local Interpretable Model-agnostic Explanations
- Let's watch a quick video:
<https://youtu.be/hUnRCxnydCc>

EXAMPLE XAI: LIME

- Local Interpretable Model-agnostic Explanations
- Let's watch a quick video:
<https://youtu.be/hUnRCxnydCc>



WHAT IS HUMAN-CENTERED XAI?

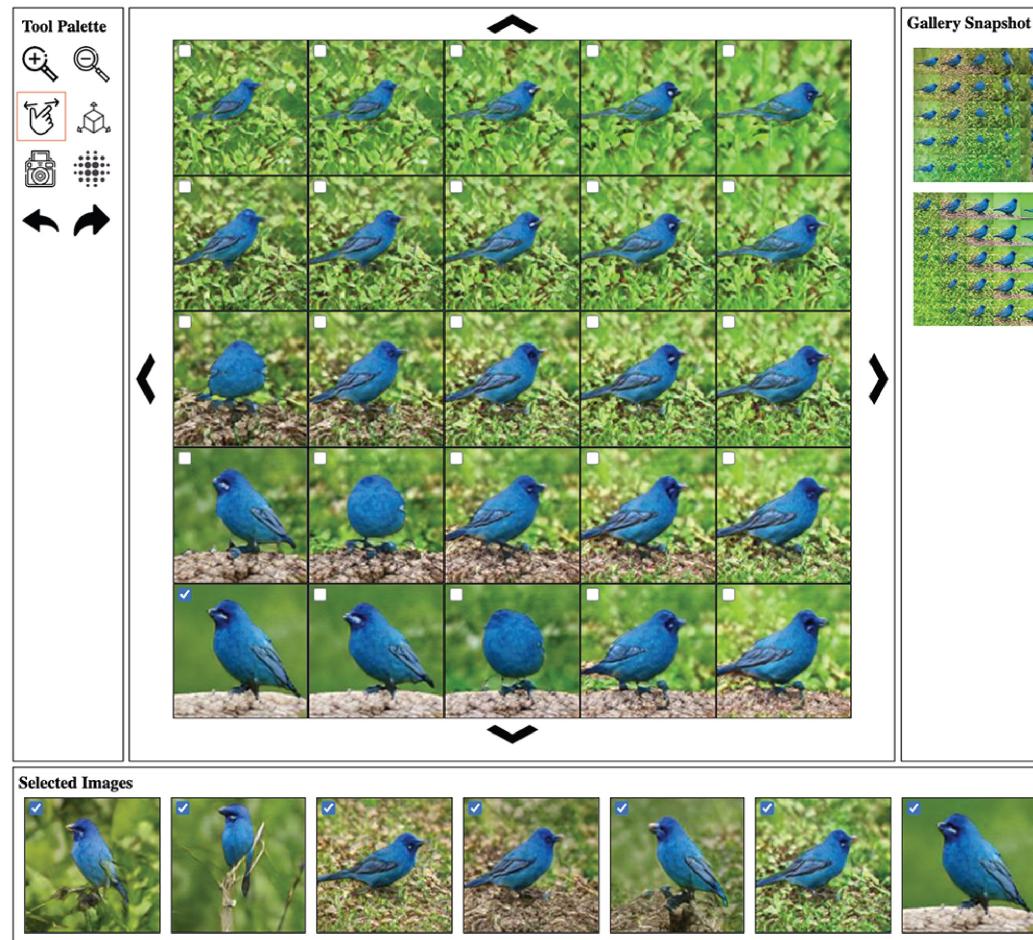
WHAT IS HUMAN-CENTERED XAI?

“XAI methods and tools
that are implemented using
a human-centered
design method.”

ALGORITHMIC AUDIT DETECTS ERRORS IN E-GOVERNMENT SYSTEMS



INTERACTIVE IMAGE GALLERIES ENABLE EXPLORATION OF GAN MODELS



Zhang & Banovic. 2021. Method for Exploring Generative Adversarial Networks (GANs) via Automatically Generated Image Galleries. In *Proc. CHI 2021*.

AUDIT BASED XAI (OR WHEN AI CANNOT BE TRUSTED TO EXPLAIN ITSELF)

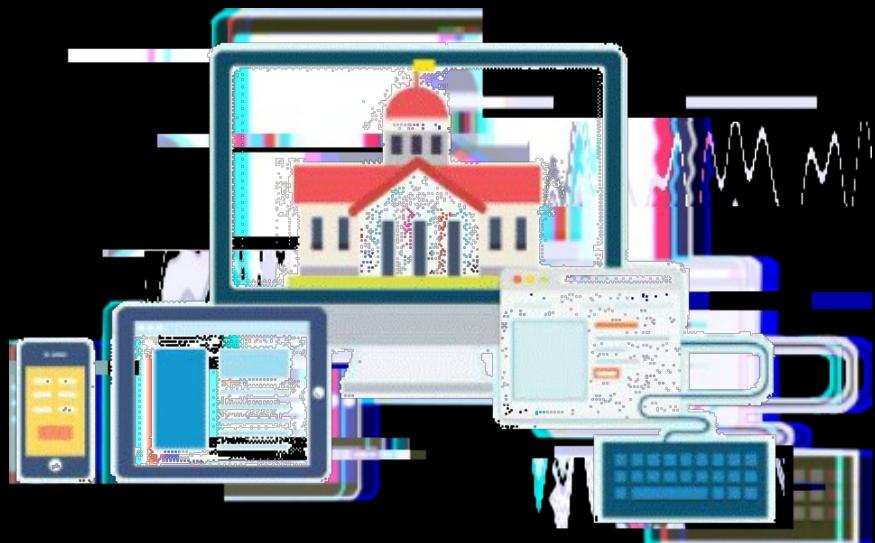
ALGORITHMIC AUDIT: E-GOVERNMENT SYSTEMS

- E-Government
 - Provision of state services through technological communication devices
 - May improve access
 - May reduce administrator workload



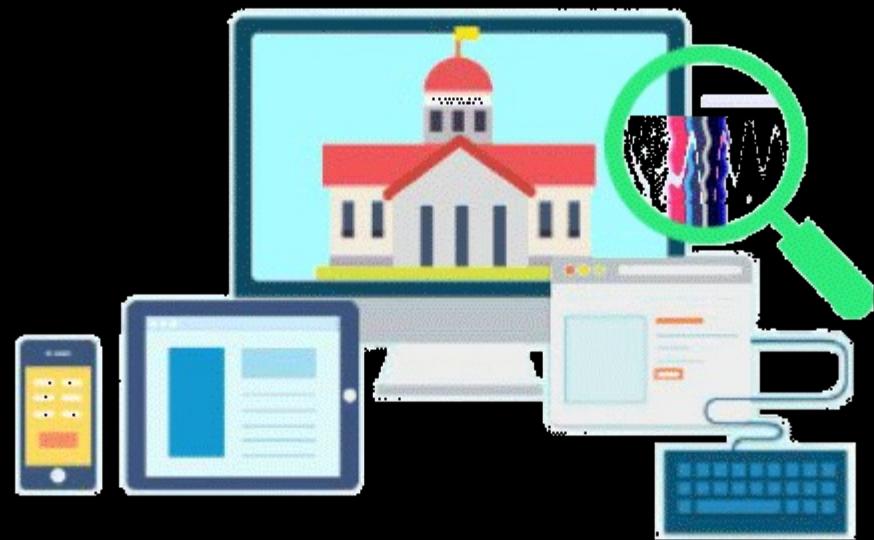
ALGORITHMIC AUDIT: E-GOVERNMENT SYSTEMS

- Huge impact for affected people
- Can be hard to detect
- Can be hard to contest



ALGORITHMIC AUDIT: E-GOVERNMENT SYSTEMS

- Detect errors before they affect people
- Make embedded errors visible
- Allow organizations and individuals to articulate demands for change



ELIGIBILITY SCREENING TOOLS

- Accepts estimated information about a household, then predicts whether they will qualify for benefits
 - Eligibility handbooks can stretch hundreds of pages, often written in complex legal language

The screenshot shows a user interface for a screening tool. At the top, it says "Am I eligible for SNAP" and "Find out if you may be eligible in 10 seconds." Below this, there's a question: "How many people live in your household, including you? (Required)". A note states: "If you buy and make more than 2/3 of your meals with others, they must be in your household. If your spouse or children under 22 live with you, they must be in your household even if you do not buy and make meals with them." There are five numbered buttons (1, 2, 3, 4, 5) and a "More" button. Next is a question: "Is anyone age 60 or older? (Required)". It has two radio buttons: "Yes" and "No". Then, "Does anyone in the household have a physical or mental disability? (Required)" with "Yes" and "No" radio buttons. Finally, "What is the total gross income for your household? (Required)" with input fields for "\$" and "per Month". At the bottom are "Go Back" and "Get my results" buttons.

ALGORITHMIC AUDIT: E-GOVERNMENT SYSTEMS

- Method for auditing Screening Tools
 - Locate embedded errors, i.e., incorrect application of the legal rules
 - Illustrated on the Pennsylvania “Do I Qualify?” screening tool

The figure consists of three side-by-side screenshots of the Pennsylvania COMPASS website. The left screenshot shows the homepage with a blue header, a central image of a doctor and patient, and two buttons: 'APPLY NOW' and 'DO I QUALIFY?'. The middle screenshot shows the 'Getting Started' step, which is part of a three-step process indicated by a progress bar at the top. It asks for details about the Head of Household, including Name, Age, and Sex, with a 'Remove' button. The right screenshot shows the 'Your Results' step, which displays a summary of potential benefits based on the answers given. It includes links for 'More Information' and 'Apply Now'.

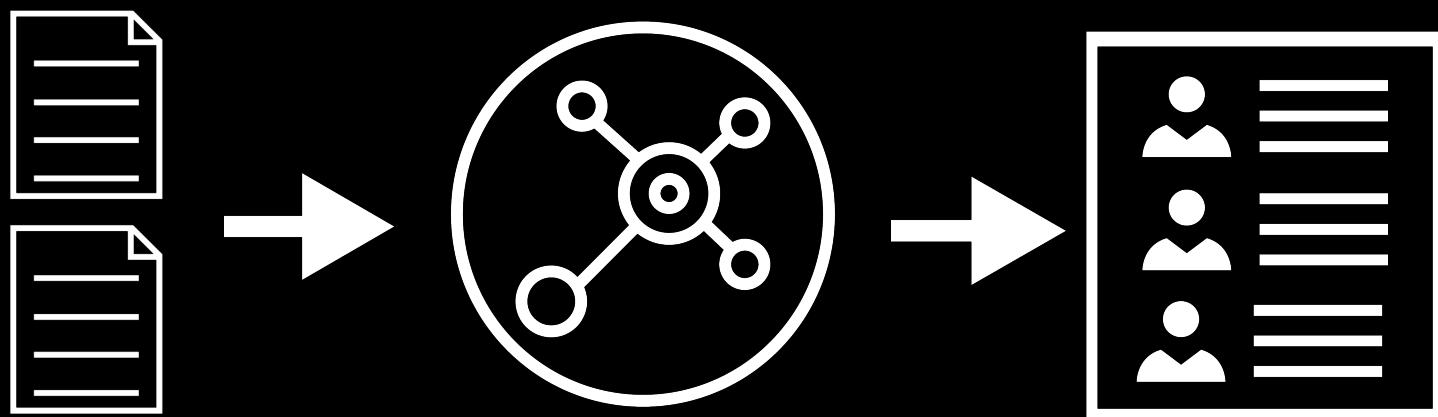
ALGORITHMIC AUDIT: E-GOVERNMENT SYSTEMS

1. Generate Test Households

External
Data Sets

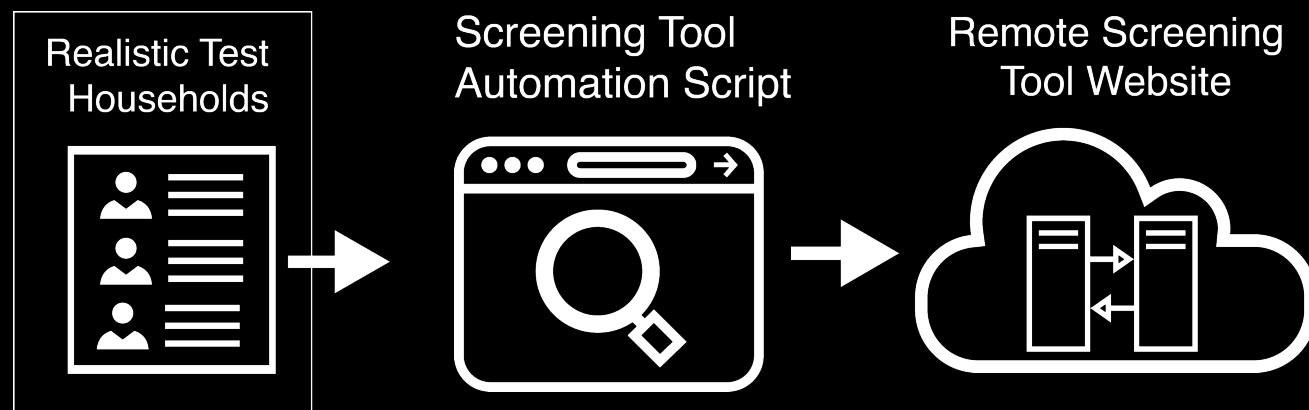
Data Transformation
and Merge Scripts

Realistic Test
Households



ALGORITHMIC AUDIT: E-GOVERNMENT SYSTEMS

2. Retrieve Test Determinations



ALGORITHMIC AUDIT: E-GOVERNMENT SYSTEMS

The screenshot shows a web browser window for the COMPASS HHS Do I Qualify website at <https://www.compass.state.pa.us/Compass.Web/Screening/DolQualify#/Household>. The page title is "Household". The interface includes a header with a green circular icon, a file icon, a home icon, and a user icon. A status bar at the top right shows "Color temperature: 6500K", "Sunset: about an hour ago, Wake: 11 hours ago", and a battery level.

Household

Please provide the details about the Head of Household first.

Name * Age * Sex * Male Female Remove

ADD ANOTHER PERSON If there is anyone else in the household, please click the 'Add Another Person' button.

About how much is the total value of all the resources owned by the people in the household? ?
Format: XXXXXXX.XX

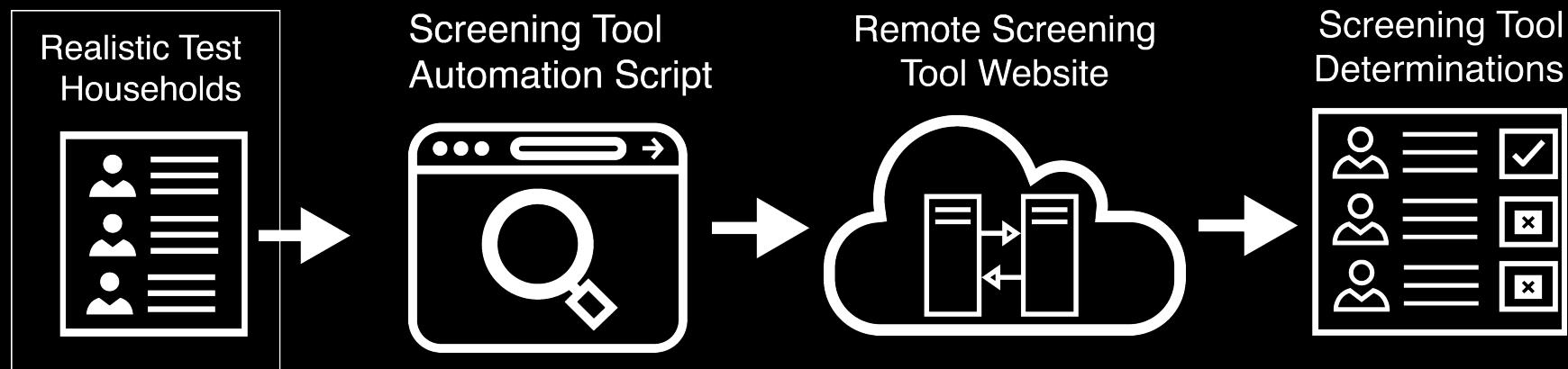
Does anyone in the household who is 21 or younger have a parent who does not live in the house or who has died?
 Yes No

Does anyone in the household have a spouse who is not living in the house or has died?
 Yes No

Has anyone in the household lost their job or had their hours reduced through no fault of their own within the past year?
 Yes No

ALGORITHMIC AUDIT: E-GOVERNMENT SYSTEMS

2. Retrieve Test Determinations



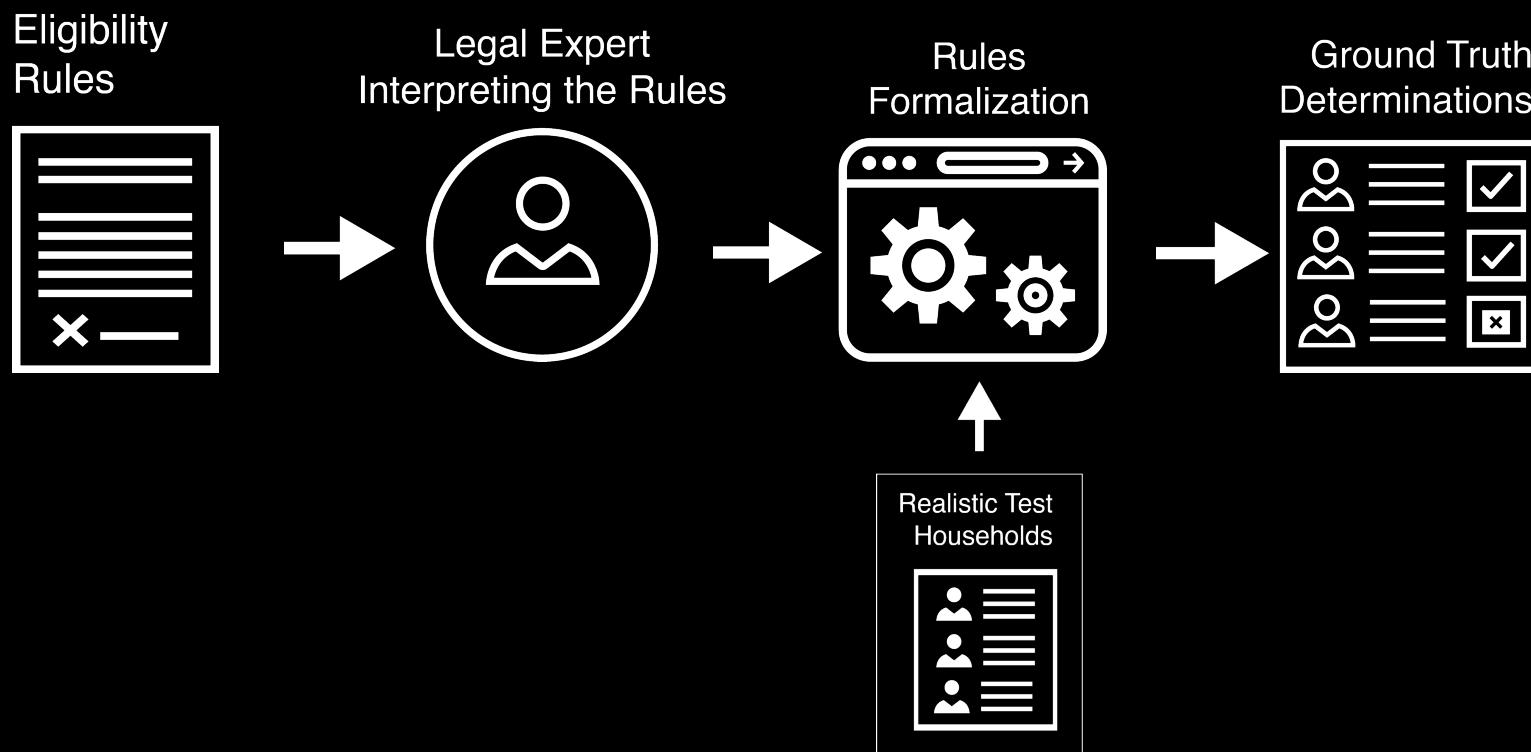
ALGORITHMIC AUDIT: E-GOVERNMENT SYSTEMS

3. Translate Legal Text to Code



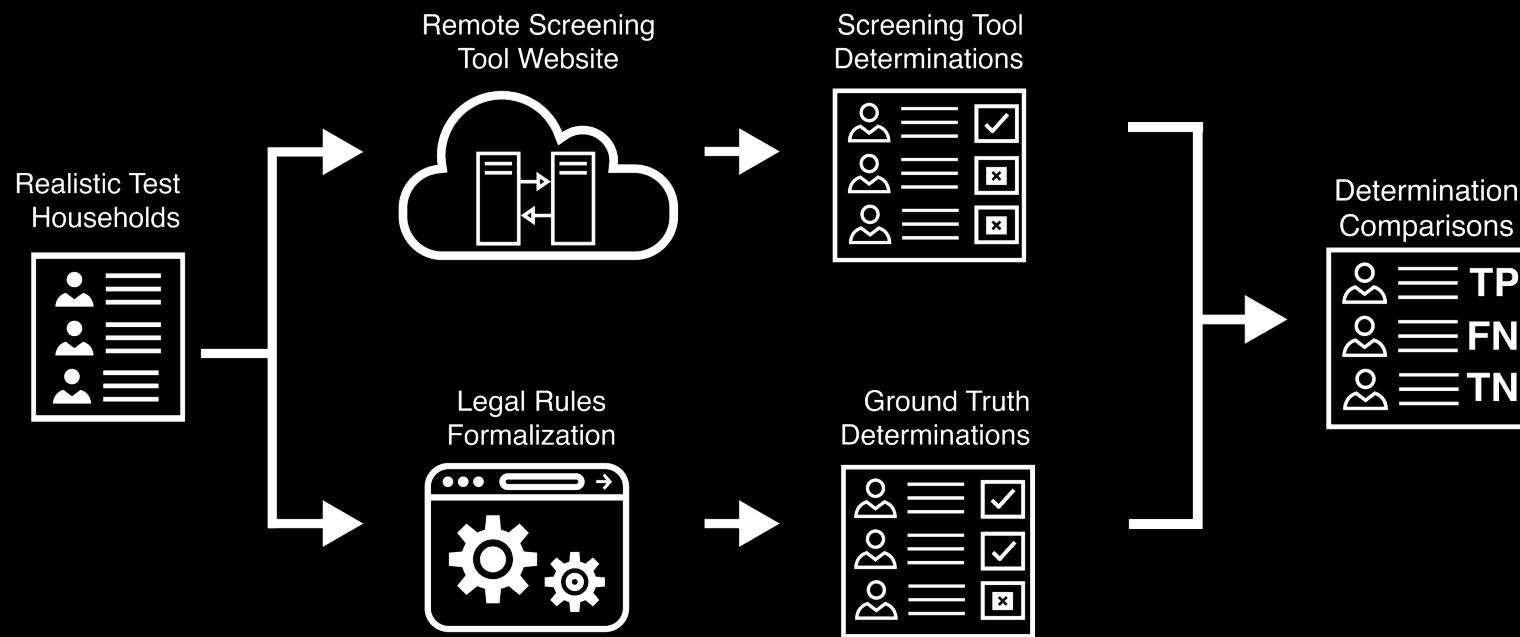
ALGORITHMIC AUDIT: E-GOVERNMENT SYSTEMS

3. Translate Legal Text to Code



ALGORITHMIC AUDIT: E-GOVERNMENT SYSTEMS

4. Identify Conflicting Determinations

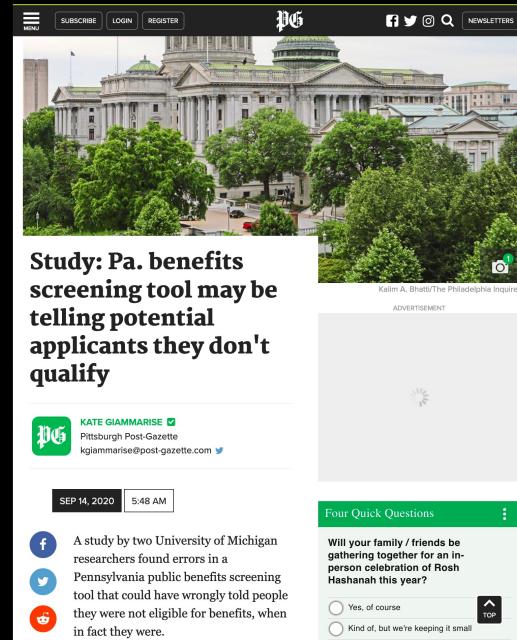


ALGORITHMIC AUDIT: E-GOVERNMENT SYSTEMS

- Example Results: Subsidized Child Care
 - Program described in Pennsylvania Code Title 55, Chapter 3041
 - Families with a child who needs supervision and whose income falls below 200% of the Federal Poverty Income Guidelines (FPIG) are eligible
 - No families predicted eligible for the subsidized child care benefit
 - Our formalization marked at least 4.6% of families as eligible

ALGORITHMIC AUDIT: E-GOVERNMENT SYSTEMS

- Some of the errors we exposed have been corrected in the “Do I Qualify?” Screening Tool
- Suggests that coordination between interested groups & public attention required to correct errors



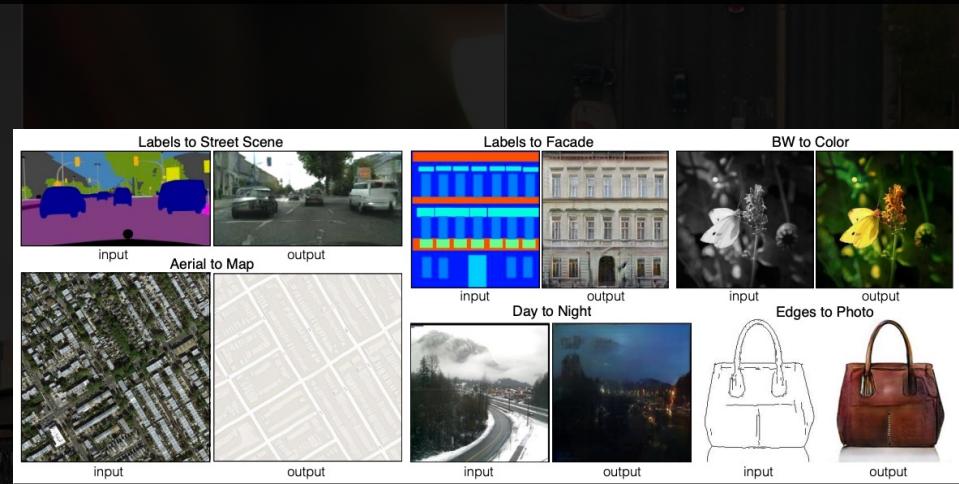
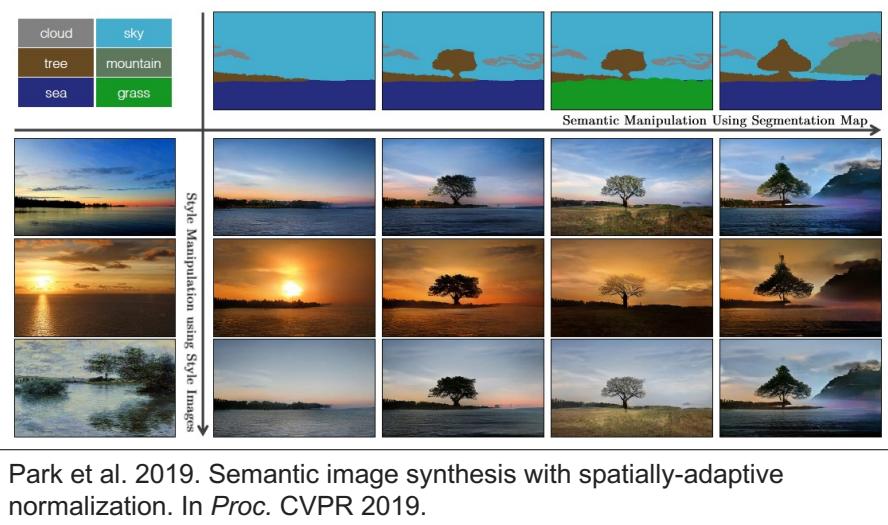
LESSONS FROM ALGORITHMIC AUDIT

- Simulating interactions of people with an opaque system can expose inner working of the system
- Algorithmic audit generates actionable trail of issues with these complex systems that can be used to fix those systems

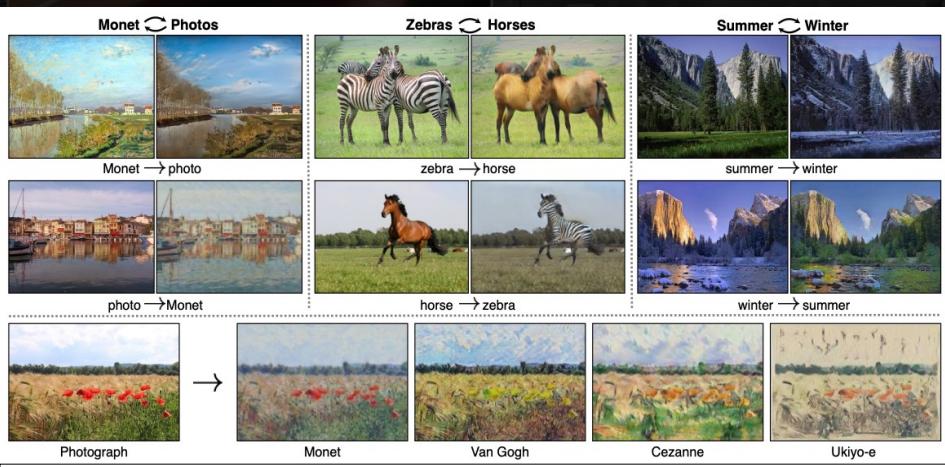
HUMAN-CENTERED XAI THROUGH INTERACTION

EXPLORATION BASED XAI INTERFACES

INTERACTING WITH GANs



Isola et al. 2017. Image-to-image translation with conditional adversarial networks. In *Proc. CVPR 2017*.



Zhu et al. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. ICCV 2017*.

INTERACTING WITH GANs

- Training a GAN involves two networks:
 1. a **Generator** that takes in a vector of latent variables $z = (z_1, \dots, z_n) \in \mathbb{R}^n$ and outputs the corresponding image
 2. a **Discriminator** that is used to distinguish between generated images and real images (training data).

INTERACTING WITH GANs

- Training a GAN involves two networks:
 1. a **Generator** that takes in a vector of latent variables $z = (z_1, \dots, z_n) \in \mathbb{R}^n$ and outputs the corresponding image
 2. a **Discriminator** that is used to distinguish between generated images and real images (training data).
- The Generator is trained to maximize the probability of fooling the Discriminator

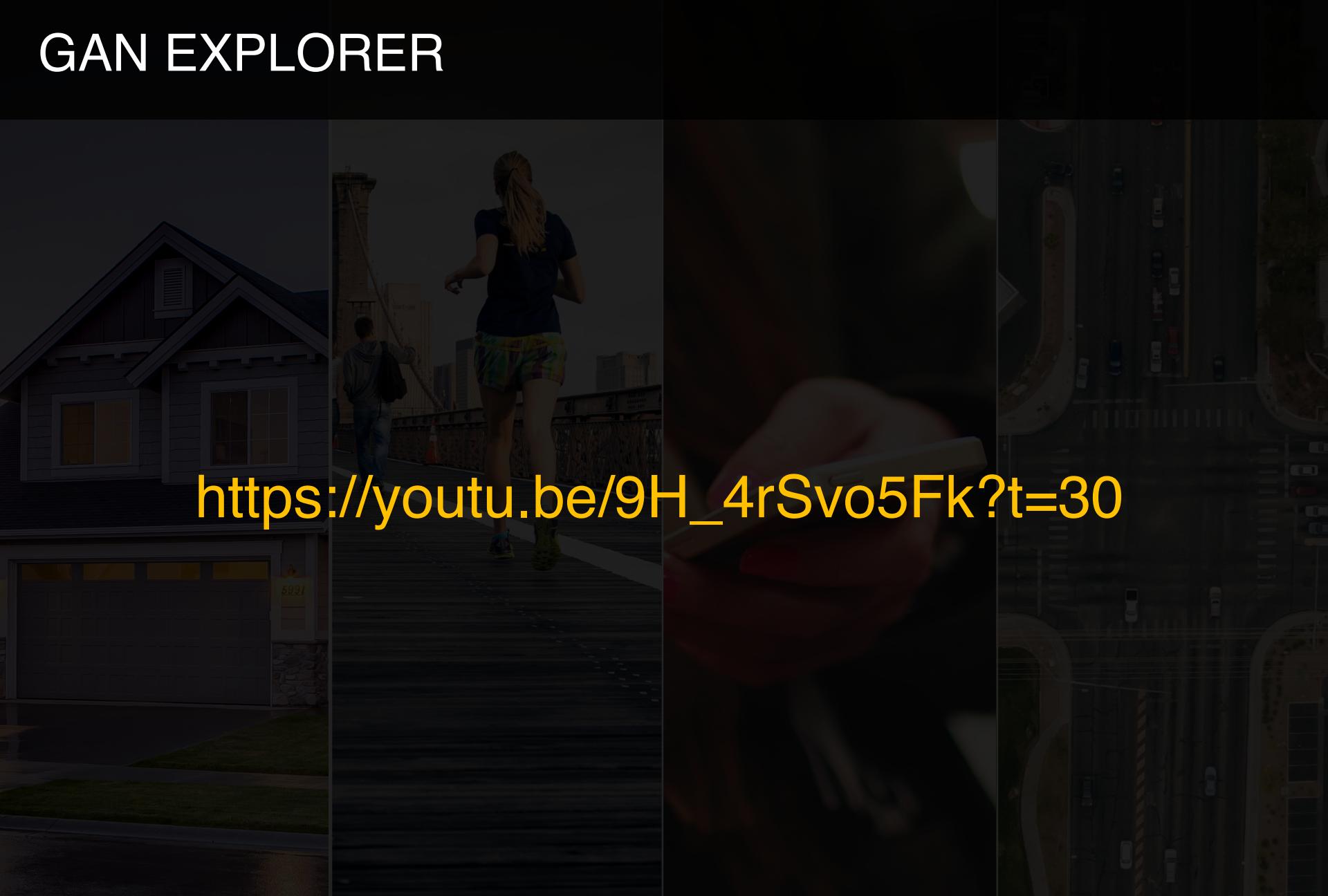
INTERACTING WITH GANs

- Training a GAN involves two networks:
 1. a **Generator** that takes in a vector of latent variables $z = (z_1, \dots, z_n) \in \mathbb{R}^n$ and outputs the corresponding image
 2. a **Discriminator** that is used to distinguish between generated images and real images (training data).
- The Generator is trained to maximize the probability of fooling the Discriminator
- The Discriminator is trained to discriminate training data from the images created by the Generator.

INTERACTING WITH GANs

- Challenging to evaluate quality of images that a GAN can generate (e.g., ability to generate a diverse set of photo-realistic images)
- Often using tedious visual examination of image galleries generated from narrow probability distributions

GAN EXPLORER



https://youtu.be/9H_4rSvo5Fk?t=30

EVALUATING GAN EXPLORER

- **Study 1:** Evaluate GAN Explorer interactions to show that users can select images using the tool

STUDY 1: EVALUATING GAN EXPLORER

- Amazon MTurk study with 367 participants
- Task: select photo-realistic images generated from the BigGAN model using our interactive tool
- Participants selected images from 10 BigGAN categories.
- We measured how many images participants selected and what tools they used to select them

STUDY 1: EVALUATING GAN EXPLORER

- Participants interacted with:

Tool	# times	Tool	# times
Pan	2,138	Randomize	679
Zoom in	969	Undo	97
Zoom out	107	Redo	15
Zoom into region	210	Take snapshot	200
Pivot	150	Revert to previous snapshot	34

- Participants selected 10,026 GAN-generated images using our prototype

EVALUATING GAN EXPLORER

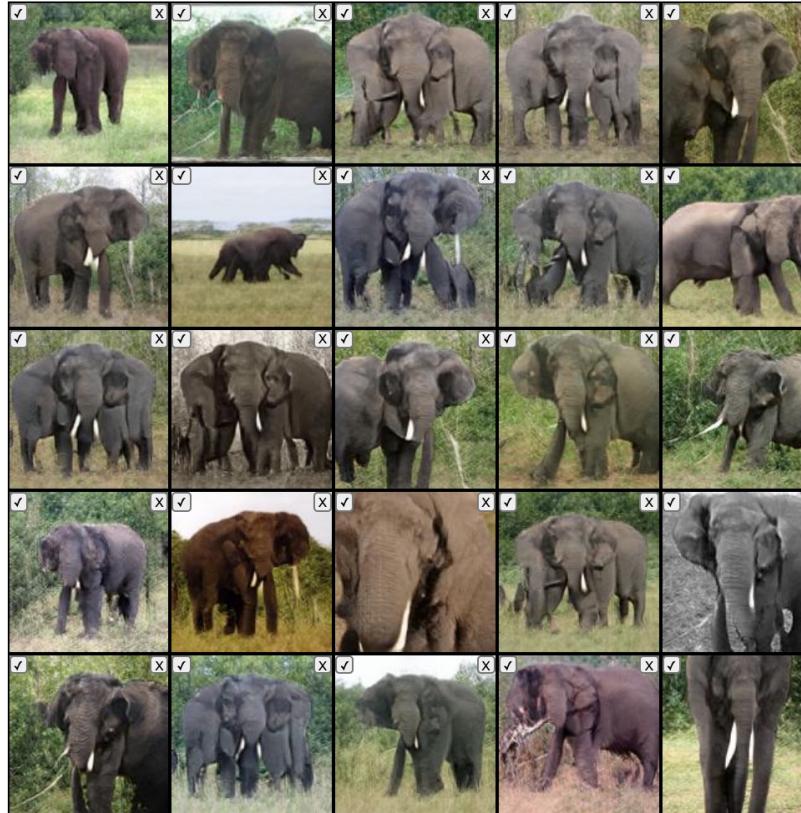
- Study 1: Evaluate GAN Explorer interactions to show that users can select images using the tool
- **Study 2:** Validate that images that the participants in Study 1 selected are photo-realistic

STUDY 2: VALIDATING IMAGE QUALITY

- Amazon MTurk study with 1,622 participants
- Task: select only photo-realistic images from an image gallery
- We measured how many images from Study 1 were photo-realistic

STUDY 2: VALIDATING IMAGE QUALITY

Please select all **highly photo-realistic** images that occur only once in the gallery below by clicking on the checkbox in the upper left corner of the image. For all other images, click on the checkbox in the upper right corner of the image. Once you have evaluated all 25 images in the gallery, press on "Submit Images" button below.



Submit Images

STUDY 2: VALIDATING IMAGE QUALITY

- Out of 10,026 images generated in Study 1, 8,015 images (79.94%) were rated as photo-realistic by at least one participant in Study 2
- We observed a lot of ambiguity in ratings, and low inter-rater agreement
- Visual examination of images rated as photo-realistic by at least 75% of participants uncovered a diverse set of photo-realistic images

EVALUATING GAN EXPLORER

- Study 1: Evaluate GAN Explorer interactions to show that users can select images using the tool
- Study 2: Validate that images that the participants in Study 1 selected are photo-realistic
- **Study 3:** Compare our image galleries with galleries sampled and generated using existing methods

AUTOMATICALLY GENERATED IMAGE GALLERIES

- Our method: Using MCMC to sample from the posterior probability distribution of latent vector z , given user-selected photo-realistic images
- Baseline: sampling the latent vector z from a truncated normal distribution

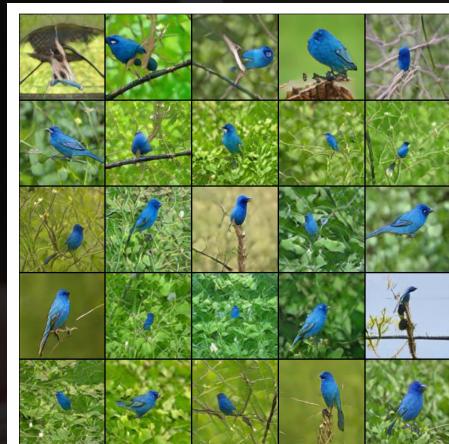
STUDY 3: EVALUATING IMAGE GALLERIES

- Amazon MTurk study with 1,000 participants
- Randomly assigned participants into one of 50 (*Method x Category*) conditions
 - *Method*: our method and four baselines
 - *Category*: one of 10 BigGAN image categories
- Task: select photo-realistic images that occur only once in the assigned gallery containing 25 images
- We measured number of photo-realistic images per *Method*

STUDY 3: EVALUATING IMAGE GALLERIES

BigGAN heuristic-based thresholds

Our method



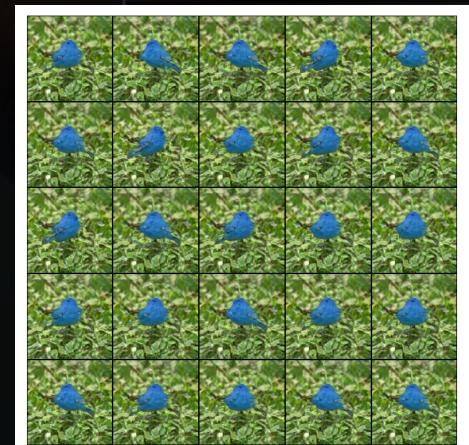
Truncated at 2



Truncated at 1



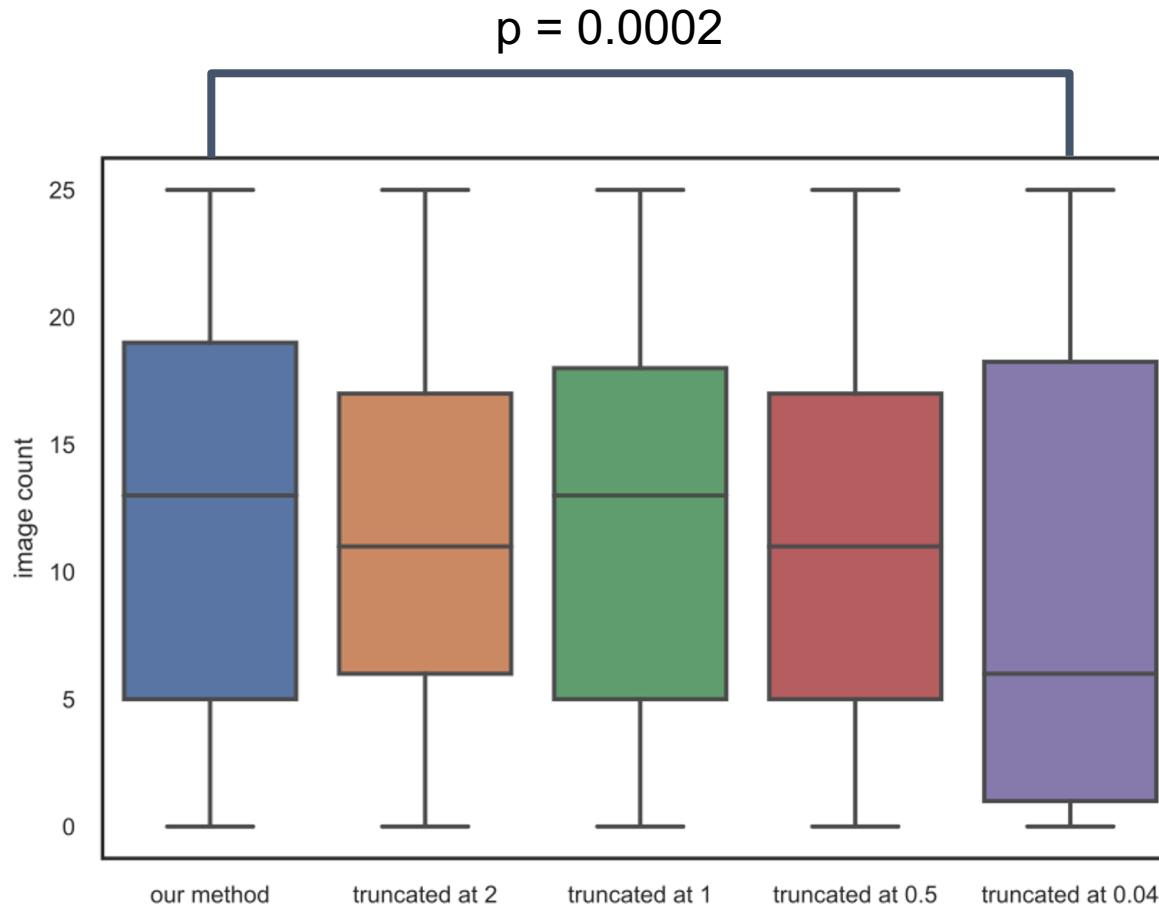
Truncated at 0.5



Truncated at 0.04

STUDY 3: EVALUATING IMAGE GALLERIES

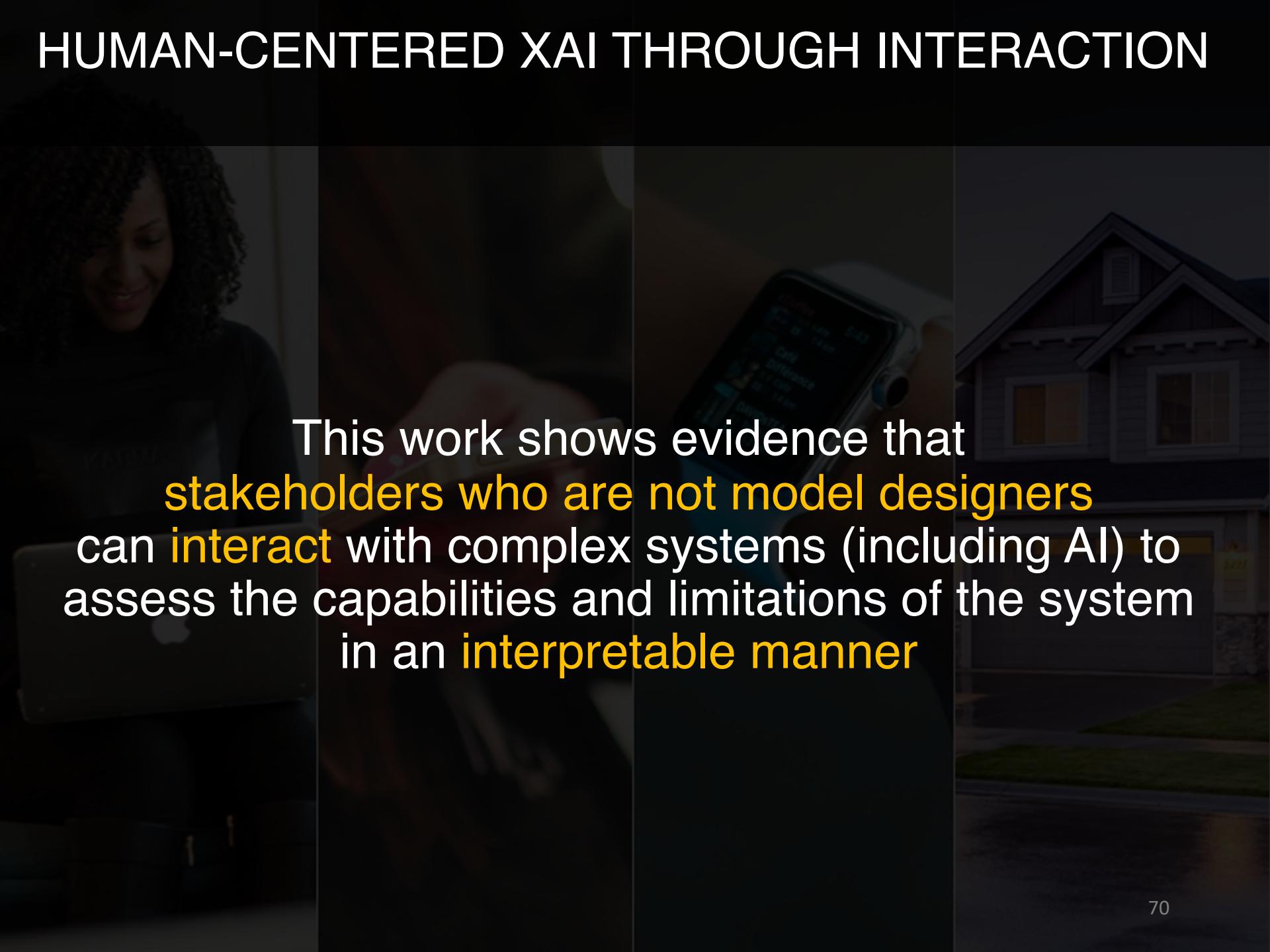
- Our tests found significant main effect of *Method* on the number of selected images ($p=0.0006$).



INTERACTING WITH GANs

- Our interactive tool enables **qualitative GAN exploration** via interactive visual examination
- Our automated image gallery generation method enables quick **creation of many diverse, photorealistic image galleries** to support interactive qualitative evaluation of GANs
- GAN Explorer demo:
<https://comphcithree.eecs.umich.edu:8040/intro/>

HUMAN-CENTERED XAI THROUGH INTERACTION



This work shows evidence that stakeholders who are not model designers can interact with complex systems (including AI) to assess the capabilities and limitations of the system in an interpretable manner

FUTURE DIRECTIONS

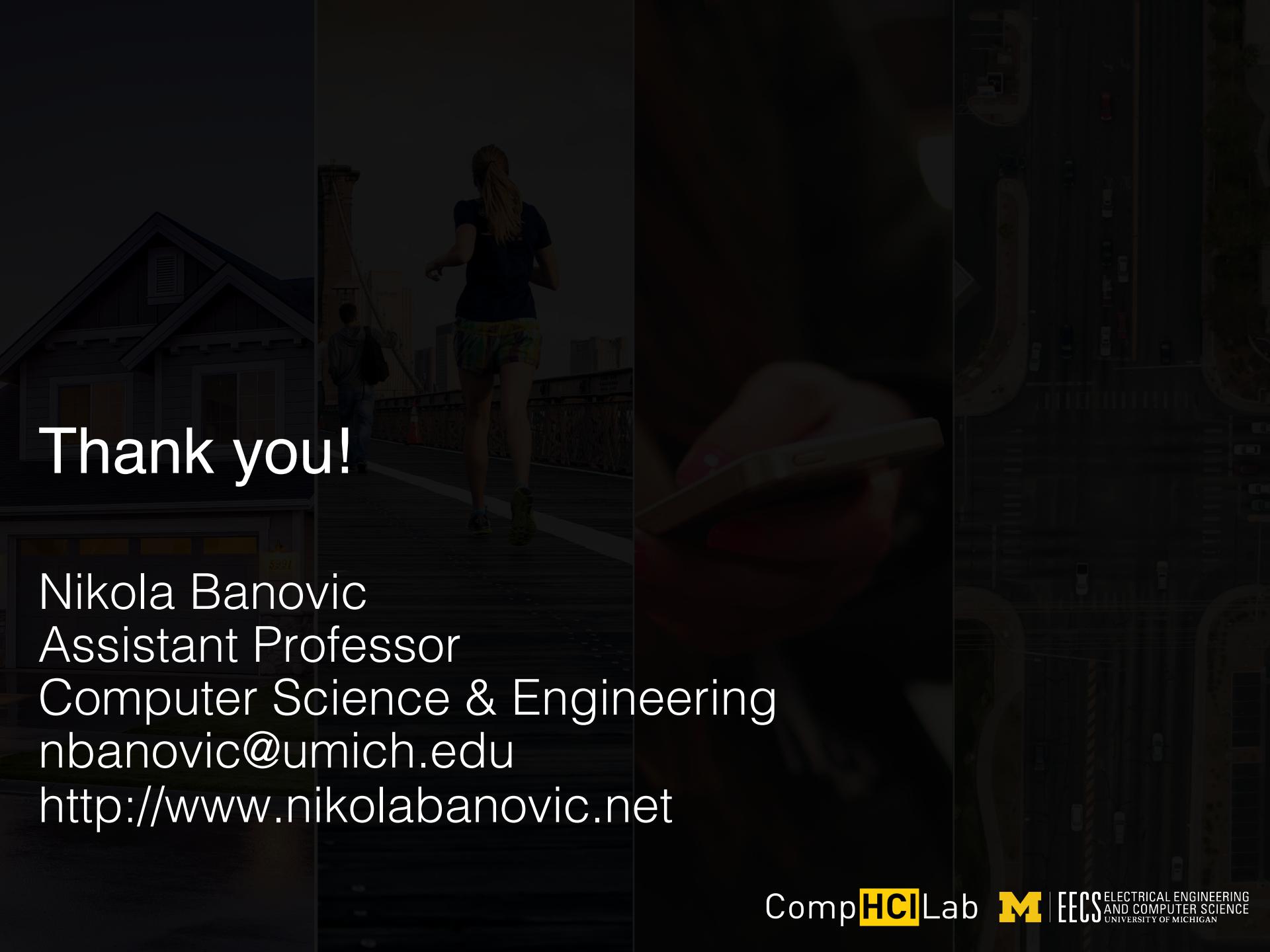
- Explore variety of interaction techniques to identify promising threads of research

FUTURE DIRECTIONS

- Explore variety of interaction techniques to identify promising threads of research
- Formulate the interactive model exploration as an optimal sequential experimental design problem

FUTURE DIRECTIONS

- Explore variety of interaction techniques to identify promising threads of research
- Formulate the interactive model exploration as an optimal sequential experimental design problem
- Go beyond identifying capabilities and limitations of models to include other factors that impact people and our society more broadly (e.g., ethics, fairness)



Thank you!

Nikola Banovic
Assistant Professor
Computer Science & Engineering
nbanovic@umich.edu
<http://www.nikolabanovic.net>

WHAT IS HUMAN-CENTERED XAI?

Next up, let's implement interactions for model exploration

INTERACTIVE & EXPLAINABLE AI

Nikola Banovic

Assistant Professor
Computer Science & Engineering
University of Michigan
nbanovic@umich.edu
<http://www.nikolabanovic.net>

The 6th Summer School on Computational Interaction – Day 4