Designing Appropriate AI for Organ Allocation and Acceptance

Casey Canfield, Associate Professor

Engineering Management & Systems Engineering

Missouri S&T

Disclosure

- This work is supported by National Science Foundation (NSF) grants (#2222801 & #2026324).

Interdisciplinary research is needed to design **appropriate** AI systems.

Develop Deep Learning Models

Identify Fairness Preferences

Support Human-AI Collaboration

Generate Evidence to Support Adoption

We have developed 3 deep learning models.

	Deceased Donor Organ Allocation (Hard-to-Place)	Provisional Acceptance (Non-Primary Offer)	Final Acceptance (Primary Offer)
Predicts	Likelihood of non-utilization for deceased donor organ	Likelihood of provisional acceptance for candidate	Likelihood of final acceptance for candidate
Use Cases for OPO	Deciding whether to pursue out-of-sequence allocation	Making open offers	Making open offers
Use Cases for Transplant Center	N/A	Setting up filters	Second opinion, facilitate decision-making for late-night offers

Initial prototype available at:
https://ddoa.mst.hekademeia.org

Ashiku, Threlkeld, Dagli, Schnitzler, Canfield, Lentine, Randall. (2022). Donor Disposition AI Model to Predict Transplant for Recovered Deceased Donor Kidneys. *American Journal of Transplantation*, 22, 652-653.

Existing biases are perpetuated in AI and deep learning models.

Training Data/Model Development

- Observed biases related to gender, diabetes, hypertension

- How can we use fairness metrics to adjust AI outputs?

Human-AI Interaction

- People over or under-rely on AI for decision-making
- Need to preserve human autonomy

- How can explainable AI (XAI) support model interpretability?

Upcoming abstracts accepted at World Transplant Congress:

- Fadaiya, Akinola, Threlkeld, Dagli. (2025). Assessment of Potential Bias in Kidney Transplant Acceptance Decision using the Deceased Donor-Recipient Matching AI Model.
- Akinola, Fadaiya, Threlkeld, Dagli. (2025). Evaluating Potential Gender Bias in Kidney Discard prediction leveraging the Deceased Donor Allocation Model.

We can improve fairness, but we must choose **one** definition.

Table 1: Classification of fairness notions.

Fairness Notion	Reference	Formulation	Classification	Type
Statistical Parity	[15]	$P(\hat{Y} \mid A=0) = P(\hat{Y} \mid A=1)$	Independence $\hat{Y} \perp A$	Group Fairness
Conditional Statistical Parity	[10]	$P(\hat{Y}=1 \mid E=e, A=0) = P(\hat{Y}=1 \mid E=e, A=1) \quad \forall e$		
Equalized Odds	[23]	$P(\hat{Y}=1 \mid Y=y,\ A=0) = P(\hat{Y}=1 \mid Y=y,\ A=1) \quad \forall y \in \{0,1\}$	Separation $\hat{Y} \perp A \mid Y$	
Equal Opportunity		$P(\hat{Y}=1 \mid Y=1, A=0) = P(\hat{Y}=1 \mid Y=1, A=1)$		
Predictive Equality	[10]	$P(\hat{Y}=1 \mid Y=0, A=0) = P(\hat{Y}=1 \mid Y=0, A=1)$		
Balance for Positive Class	[29]	$E[S \mid Y-1, A-0)] - E[S \mid Y-1, A-1]$		
Balance for Negative Class		$E[S \mid Y=0, A=0] = E[S \mid Y=0, A=1]$		
Conditional Use Accuracy Equality	[6]	$P(Y=y \mid \hat{Y}=y, A=0) = P(Y=y \mid \hat{Y}=y, A=1) \quad \forall y \in \{0,1\}$	Sufficiency $Y \perp A \mid \hat{Y}$	
Predictive Parity	[8]	$P(Y=1 \mid \hat{Y}=1, A=0) = P(Y=1 \mid \hat{Y}=1, A=1)$		
Calibration		$P(Y=1 \mid S=s, A=0) = P(Y=1 \mid S=s, A=1) \quad \forall s \in [0,1]$		
Well-calibration	[29]	$P(Y=1 \mid S=s, A=0) = P(Y=1 \mid S=s, A=1) = s \quad \forall\, s \in [0,1]$		
Overall Accuracy Equality	[6]	$P(\hat{Y}=Y \mid A=0) = P(\hat{Y}=Y \mid A=1)$	Other metrics from confusion matrix	
Treatment Equality		$\frac{FN}{FP}(\mathrm{a=0}) = \frac{FN}{FP}(\mathrm{a=1})$		
Total Fairness		-	Independence, Separation and Sufficiency	
No unresolved discrimination	[27]	-	Causality	
No proxy discrimination		$P(\hat{Y} \mid do(P_x = p)) = P(\hat{Y} \mid do(P_x = p\prime)) \quad \forall P_x \quad and \quad \forall\, p, p\prime$		
Counterfactual Fairness	[30]	$P(\hat{Y}_{A\leftarrow a}(U) = y \mid X=x, A=a) = P(\hat{Y}_{A\leftarrow a\prime}(U) = y \mid X=x, A=a)$		Individual Fairness
Causal Discrimination	[20]	$X_{(\mathrm{a=0})} = X_{(\mathrm{a=1})} \wedge A_{(\mathrm{a=0})} \neq A_{(\mathrm{a=1})} \Rightarrow \hat{y}_{(\mathrm{a=0})} = \hat{y}_{(\mathrm{a=1})}$	Similarity Metric	
Fairness Through Awareness	[15]	$D(M(v_i), M(v_j)) \leq d(v_i, v_j)$		

Makhlouf, Zhioua, Palamidessi. (2021). On the Applicability of Machine Learning Fairness Notions. *SIGKDD Explor. Newsl.*, 23(1), 14–23.

Disagreement feedback can be used to measure fairness preferences.

DONOR

Age	**28**
Race	White
Gender	Male
Kidney Quality	18

Recipient #	1	7	8	9	10
Age	19	67	71	73	38
Race	White	White	White	White	Multi-Racial
Gender	Male	Male	Male	Male	Female
Est. Post Transplant Survival	3	67	61	65	17
Distance from Transplant Center	0.0 miles	65.0 miles	0.0 miles	0.0 miles	199.0 miles
Acceptance Rate	59%	78%	78%	78%	59%
Surgeon's Decision	No Transplant	Transplant	No Transplant	No Transplant	Transplant

Next Step:

Collect input from transplant stakeholders (patients, donors, OPOs, transplant centers)

Given the **surgeon's decision** (highlighted in purple), what is the fairness of the Acceptance Rates from the AI recommender system?

Completely Unfair	Moderately Unfair	Slightly Unfair	Neither Fair nor Unfair	Slightly Fair	Moderately Fair	Completely Fair
○	○	○	○	○	○	○

Telukunta, Nadendla. (2023). Towards Inclusive Fairness Evaluation via Eliciting Disagreement Feedback from Non-expert Stakeholders. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. 245-260.

For human-AI collaboration, the value of AI varies by person and by case.

OPOs

- Desired an immediate intervention to **screen** for a hard-to-place kidney.

- Preference for XAI information to always be reviewed

Transplant Centers

- Transplant centers framed the AI as a **second opinion** to accept or deny a kidney offer.

- Value of AI depends on the difficulty of case

Subramanian, Canfield, Shank. (2024). Designing explainable AI to improve human-AI team performance: a medical stakeholder-driven scoping review. *Artificial Intelligence in Medicine*, *149*, 102780.
Shank. (2025). *The Machine Penalty: The Consequences of Seeing Artificial Intelligence as Less Than Human*. Springer Nature.

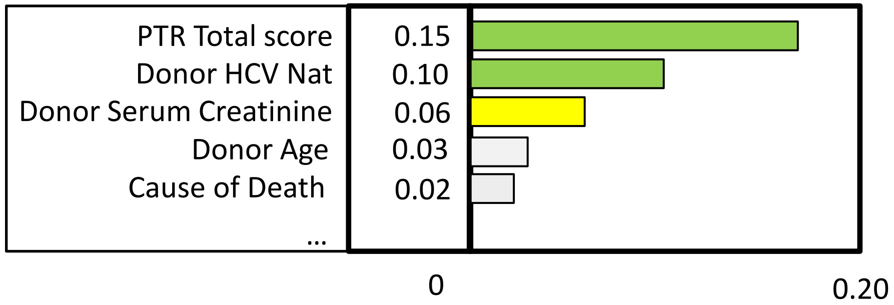
System-level XAI helps users build a mental model of how the AI works.

Examples of System-level XAI

Model Card

Training Dataset Information	
Dataset Timeline	2016-2021
No. of observations	1,300,000
No. of features	29
…	…

Feature Importance



Initial Experimental Findings

- People can use AI limitations
- But they are still inclined to believe an AI is capable outside of the training boundary

HomeVise AI

- Model Metrics
  - Accuracy = 83%
- Limitations
  - The AI's training data underrepresents buyers aged 45 and below.
  - The AI's training data underrepresents Black and Asian buyers.

Subramanian, Canfield, Shank. (in prep).

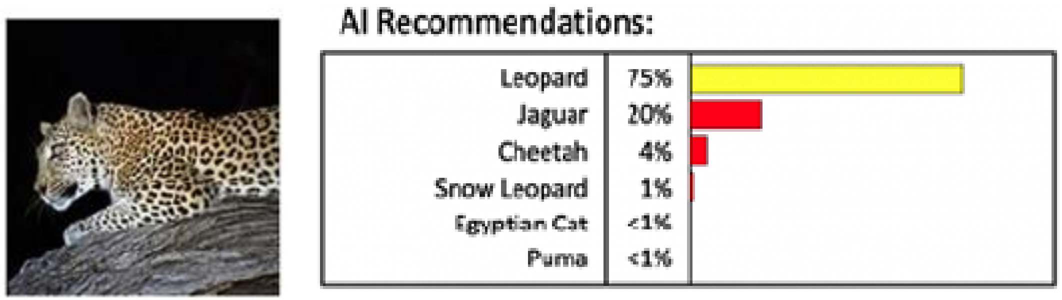
Prediction-level XAI is influenced by both the interface and decision context.

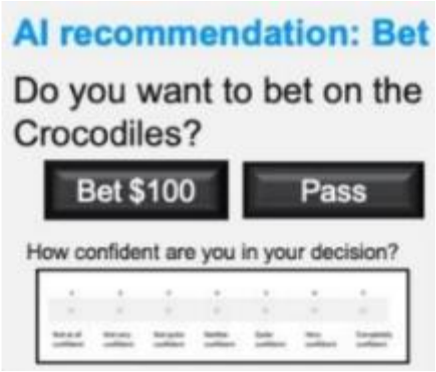
Uncertainty Information

- People can appropriately leverage uncertainty information based on their domain knowledge

Risky Context

- Recommendations to not act (pass) matter more
- People trust reliable AIs more





Subramanian, Canfield, Shank, Kinnison. (2023). Combining uncertainty information with AI recommendations supports calibration with domain knowledge. *Journal of Risk Research*, *26*(10), 1137-1152.

Elder, Canfield, Shank, Rieger, Hines. (2024). Knowing when to pass: the effect of AI reliability in risky decision contexts. *Human Factors*, *66*(2), 348-362.

AI has potential … if developed appropriately.

- Interviews with OPO leaders suggest:
  - There is hope AI will alleviate bias concerns
    - “[The transplant system is] loaded with bias. […] Ideally, the algorithms will not be loaded up with so much of that bias.”
  - AI benefits may extend beyond performance/accuracy
    - “It would ideally reduce that amount of time.”
    - “I think it'll save on staff fatigue.”

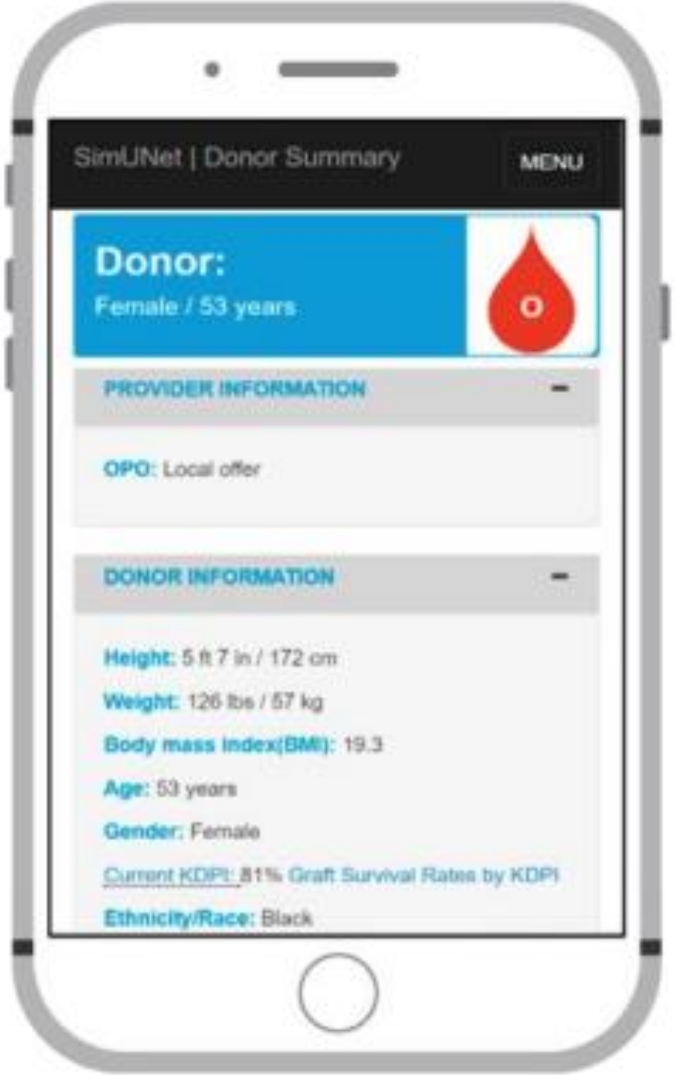
- AI developers need to be addressing issues related to bias directly by:
  - Evaluating observed bias in AI outputs
  - Conducting field trials to evaluate human-AI interaction effects

Babaee, Shank, Canfield. (in prep).

This is all building towards a field trial of our AI tool.

- SimUNet is a behavioral science platform developed by UNOS
- Mimics the real DonorNet mobile platform

- In 2026, we will conduct a Randomized Controlled Trial of the AI tool to evaluate effect on:
    - Task performance
    - Efficiency



Stewart, Shepard, Rosendale, McGehee, Hall, Gupta, ... Klassen. (2020). Can behavioral research improve transplant decision-making? A mock offer study on the role of kidney procurement biopsies. *Kidney360*, *1*(1), 36-47.

Thank you!

For more information:
https://sites.mst.edu/aifortransplant/

MISSOURI S&T


Casey Canfield (PI)
Eng Mgt & Sys Eng


Cihan Dagli (Co-PI)
Eng Mgt & Sys Eng


Daniel Shank (Co-PI)
Psych


Sid Nadendla
Computer Science

Lirim Ashiku, PhD
Richard Threlkeld, PhD
Hannah Elder, MS
Elham Babaee, MS
Hari Subramanian, MS
Rachel Dzieran
Mukund Telukunta
Matthew Kinnison

Sukruth Rao
Gabriella Stickney
Grace Hall
Aiden Pickett
Stephanie Rose
Donna Murphey
Charles Love

SAINT LOUIS UNIVERSITY


Mark Schnitzler (Co-PI)
Health Economics


Krista Lentine (Co-PI)
Nephrology


Henry Randall
Transplant Surgery


Jason Eberl
Health Care Ethics


Michael Miller
Health Care Ethics

UNOS
UNITED NETWORK FOR ORGAN SHARING


Laura Cartwright
Behavioral Science


Samantha Noreen
Research Science


Brendon Cummiskey
Behavioral Science

Join our Mailing List to stay updated!