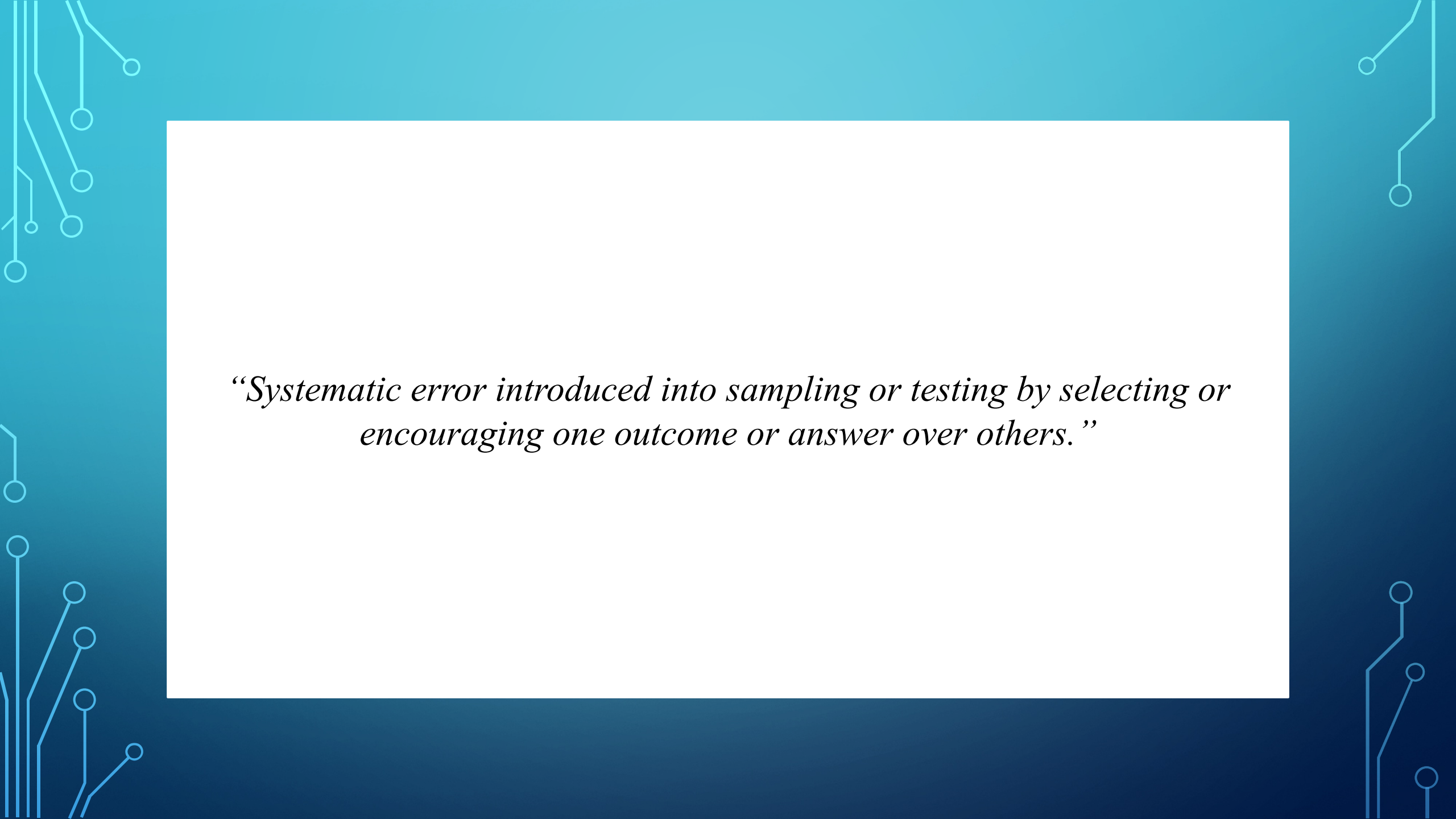


BIAS REDUCTION IN MACHINE TRANSLATION

“Autonomous intelligent systems should be designed and constructed so:

- 1. the data they are trained on is diverse, inclusive, and free from undesirable biases,*
- 2. they use that data in ethical ways,*
- 3. the decisions they make can be traced back and understood by human beings.”*

Winfield, A. F., Michael, K., Pitt, J., & Evers, V. (2019). Machine ethics: The design and governance of ethical AI and autonomous systems. *Proceedings of the IEEE*, 107(3), 509–517

The background of the slide is a gradient of blue, transitioning from a lighter shade at the top to a darker shade at the bottom. In the corners, there are decorative white line art elements that resemble electronic circuit boards, with lines and small circles representing components.

“Systematic error introduced into sampling or testing by selecting or encouraging one outcome or answer over others.”

A decorative background featuring light blue circuit-like lines and nodes on a dark blue gradient. The lines are stylized, resembling a network or data flow, with some nodes highlighted as small circles.

1. Data sets quality = performance of AI systems

Data may contain socially constructed biases, inaccuracies, errors and mistakes

Needs to be addressed prior to training with any given data set

2. The integrity of the data must be ensured

Feeding malicious data into an AI system may change self-learning systems' behaviour

3. Processes and data sets used must be tested and documented at each step


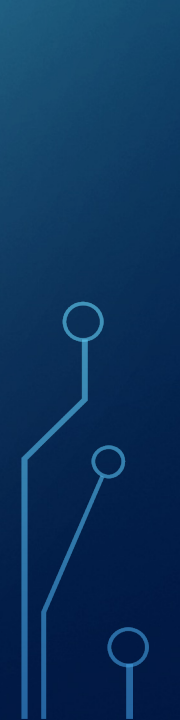
This should also apply to not in-house developed, acquired AI

The background is a dark blue gradient. In the corners, there are white line-art illustrations of circuit boards or neural network connections, featuring lines and small circles.

Problem: How to identify bias in the first place?



Solution: Techniques to avoid and mitigate unconscious bias in our data:

1. Identify human biases (= we grew up with them and have them in us)
 2. Make sure your team is diverse and follow up with diversity and anti-bias training
 3. Make sure your data is diverse and create inclusive training sets
 4. Be transparent about your algorithms and your data sources
 5. Monitor and audit your output closely and adapt your algorithm accordingly
 6. Develop a guideline on how to deal with a diversity lack in data
- 
- 

- Tomalin, M., Byrne, B., Concannon, S. *et al.* **”The practical ethics of bias reduction in machine translation: why domain adaptation is better than data debiasing.”** *Ethics Inf Technol* (2021). <https://doi.org/10.1007/s10676-021-09583-1>
- Mehrabi, Ninareh & Morstatter, Fred & Saxena, Nripsuta & Lerman, Kristina & Galstyan, Aram (2019) In-depth **“Survey on Bias and Fairness in Machine Learning”** with many insights, definitions and examples
- **“Coded Bias”, a Netflix documentary** investigates the bias in algorithms after M.I.T. Media Lab researcher Joy Buolamwini uncovered flaws in facial recognition technology.
- **EU Ethics guidelines for trustworthy AI**
- **Tools to reduce bias in AI:** AI Fairness 360 and Watson OpenScale (IBM), What-If Tool (Google)



SELMA

Stream Learning for Multilingual Knowledge Transfer

<https://selma-project.eu/>
@SELMA_project