



COMMUNICATING AI TO POLICYMAKERS AND NON- TECHNICAL DECISION-MAKERS

**PRESENTED BY
VIVI CECILIE SVENDSEN**

**A SYSTEM CAN BE EXPLAINED
AND STILL BE MISUNDERSTOOD**

A SYSTEM CAN BE **EXPLAINED** AND STILL BE **MISUNDERSTOOD**

WHAT THEY COULD HAVE HEARD:

**“A RISK MODEL HELPS THE TAX
AUTHORITY IDENTIFY APPLICATIONS
THAT MAY REQUIRE ADDITIONAL REVIEW.”**

EXTOLLED VIRTUES:

**EFFICIENCY. PRIORITIZATION.
DECISION SUPPORT. HUMAN REVIEW.**



A SYSTEM CAN BE **EXPLAINED** AND STILL BE **MISUNDERSTOOD**

WHAT THEY NEEDED TO UNDERSTAND:

- WHAT MAKES SOMEONE **“HIGH RISK”**?
- WHAT HAPPENS **AFTER THEY ARE FLAGGED?**
- **HOW MUCH** DOES THE FLAG **SHAPE** THE HUMAN DECISION?
- **WHO** IS **DISPROPORTIONATELY** FLAGGED?
- **WHO CARRIES THE COST OF A FALSE POSITIVE?**
- **CAN THEY CHALLENGE IT BEFORE THE HARM IS DONE?**



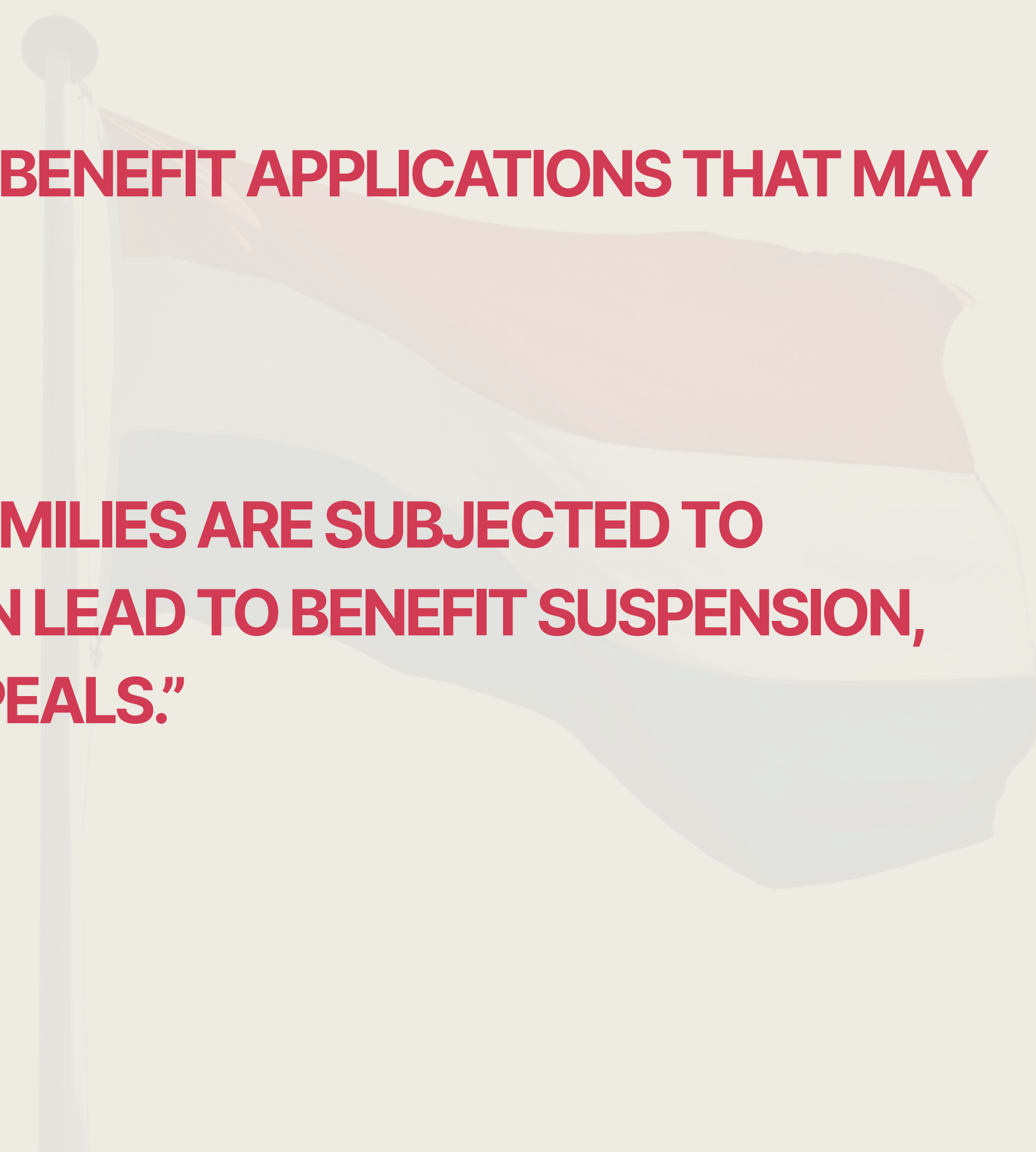
A SYSTEM CAN BE **EXPLAINED** AND STILL BE **MISUNDERSTOOD**

TECHNICALLY ACCURATE:

“THE SYSTEM HELPS IDENTIFY CHILDCARE-BENEFIT APPLICATIONS THAT MAY BE HIGH RISK.”

DECISION-RELEVANT:

“THE SYSTEM HELPS DETERMINE WHICH FAMILIES ARE SUBJECTED TO HEIGHTENED FRAUD SCRUTINY, WHICH CAN LEAD TO BENEFIT SUSPENSION, REPAYMENT DEMANDS, AND YEARS OF APPEALS.”



WE ARE OFTEN EXPLAINING
THE WRONG THING

THE GOAL IS **NOT TO MAKE THE
POLICYMAKER MORE
TECHNICAL.**

**THE GOAL IS TO MAKE THE
DECISION **MORE LEGIBLE.****

TRANSLATION $\neq$ SIMPLIFICATION

SIMPLIFICATION ASKS:

HOW CAN I MAKE THIS EASIER TO UNDERSTAND?

DECISION TRANSLATION ASKS:

**WHAT DOES THIS PERSON NEED TO UNDERSTAND TO
MAKE A RELEVANT AND INFORMED DECISION?**

WHAT DOES **DECISION-RELEVANT** **TRANSLATION** LOOK LIKE?

ARTICLE 6 CONSULTATION (POLICY LANGUAGE):
“**IS THIS A SAFETY COMPONENT?**”

DECISION LANGUAGE:
“**IF THIS AI FAILS, CAN SOMEBODY GET
HURT?**”



POLICYMAKERS NEED CONSEQUENCES, NOT JUST CONCEPTS

- 1. THE METRIC → THE CONSEQUENCE**
- 2. THE SAFEGUARD → THE POWER**
- 3. THE AVERAGE → THE DISTRIBUTION**

THE METRIC → THE CONSEQUENCE

TECHNICAL LANGUAGE:

“5% FALSE-NEGATIVE RATE”



DECISION LANGUAGE

“1 IN 20 CASES LABELLED SAFE MAY ACTUALLY BE UNSAFE.”

SAME NUMBER. DIFFERENT DECISION USEFULNESS.

THE SAFEGUARD → THE POWER

“THERE IS A HUMAN IN THE LOOP.”

WHAT PEOPLE HEAR:

A HUMAN IS IN CONTROL.

WHAT WE NEED TO KNOW:

- **CAN THE HUMAN ACTUALLY SAY NO?**
- **DO THEY HAVE THE INFORMATION?**
- **DO THEY HAVE THE TIME?**
- **DO THEY HAVE THE AUTHORITY?**
- **DO WE KNOW WHETHER INTERVENTION HAPPENS IN PRACTICE?**

THE AVERAGE → THE DISTRIBUTION

“THE SYSTEM IS HIGHLY ACCURATE.”
FOR WHOM?

THE PATTERN IS THE SAME:

METRIC → WHAT DOES THE ERROR MEAN?

SAFEGUARD → WHAT POWER DOES IT ACTUALLY PROVIDE?

AVERAGE → WHO CARRIES THE FAILURE?

**WE NEED TO TRANSLATE TECHNICAL FACTS INTO
DECISION CONSEQUENCES.**

FIVE QUESTIONS TO ASK BEFORE COMMUNICATING AI:

- ◆ **DECISION:** WHAT DOES IT ACTUALLY INFLUENCE? → RANKS JOB APPLICANTS FOR INTERVIEW.
- + **CONSEQUENCE:** WHAT HAPPENS IF IT IS WRONG? → A QUALIFIED APPLICANT NEVER REACHES A HUMAN REVIEWER
- ◆ **DISTRIBUTION:** WHO CARRIES THE ERRORS? → FALSE REJECTIONS ARE CONCENTRATED IN ONE GROUP
- **UNCERTAINTY:** WHAT DON'T WE KNOW YET? → PERFORMANCE HAS NOT BEEN TESTED ON THIS POPULATION
- ▲ **INTERVENTION:** WHO CAN ACT, AND HOW? → A REVIEWER CAN OVERRIDE THE SCORE AND THE AGENCY CAN PAUSE USE.

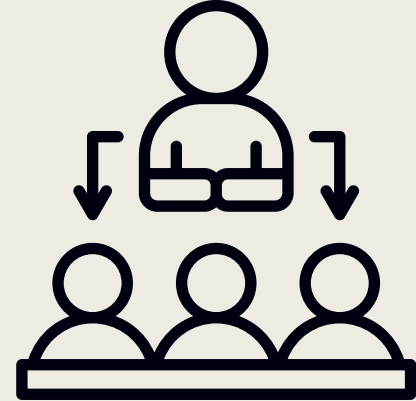
SAME TECHNICAL FACT. FOUR **DIFFERENT** DECISIONS.

TECHNICAL FACT:

**“THE SYSTEM HAS AN 8% FALSE-NEGATIVE RATE, INCLUDES HUMAN ESCALATION,
AND MAY DEGRADE AS CONDITIONS CHANGE.”**



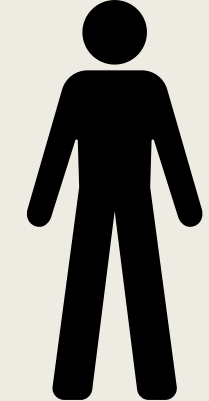
LEGISLATOR:
WHAT OBLIGATION
OR THRESHOLD
SHOULD EXIST?



EXECUTIVE:
WHAT OPERATIONAL
RISK ARE WE
ACCEPTING?



CIVIL SERVANT:
WHAT RECORDS,
OWNERSHIP, AND
REVIEW ARE REQUIRED?



AFFECTED PERSON:
WHAT HAPPENED TO ME,
WHY, AND WHAT CAN I
DO?

TRANSLATION $\neq$ SIMPLIFICATION

SIMPLIFICATION ASKS:

HOW CAN I MAKE THIS EASIER TO UNDERSTAND?

DECISION TRANSLATION ASKS:

**WHAT DOES THIS PERSON NEED TO UNDERSTAND TO
MAKE A RELEVANT AND INFORMED DECISION?**

TRANSLATION IS A **VITAL** POLICY METHOD.

PARAGRAPH → PROBLEM → FIX

PARAGRAPH:

WHAT EXACTLY DOES IT SAY?

PROBLEM:

WHAT BREAKS, BECOMES AMBIGUOUS, OR CREATES THE WRONG DECISION?

FIX:

WHAT CONCRETE CHANGE WOULD MAKE IT USABLE?



INFORMATION IS NOT UNDERSTANDING.

UNDERSTANDING IS NOT POWER TO ACT.

COMMUNICATION SHOULD CONNECT THE TWO.

MAKE THE DECISION LEGIBLE ENOUGH TO GOVERN.

EXPLAIN THE MACHINE WHEN IT MATTERS.

EXPLAIN THE **DECISION EVERY TIME.**