Promises and Pitfalls of the use of AI by the state:
A perspective from political science and theories of democracy

Keynote as part of the SICSS Summerschool

Prof. Georg Wenzelburger
Political Science | Comparative European Politics

Preliminaries I

This keynote draws on **several joint publications** appeared in the last 5 years…
My presentation **is as much my work as the work of my colleagues** both from political science, psychology, law, ethics and informatics. For references on the slides, see:

Wenzelburger, Georg/König, Pascal D./Felfeli, Julia/Achtziger, Anja (2024): "Algorithms in the Public Sector. Why context matters", in: Public Administration, 102:1, 40-60.

König. Pascal D./Felfeli, Julia/Achtziger, Anja/Wenzelburger, Georg (2024): „The importance of effectiveness versus transparency and stakeholder involvement in citizens' perception of public sector algorithms", in: Public Management Review 26:4, 1061-1082.

Wenzelburger, Georg/König, Pascal D. (2022): „The Liberal Dream of Smart Detention? Algorithms and the Politics of Pretrial Detention in the US States", in: Law & Policy, 44, 325–347.

Haeri, Maryam A./Hartmann, Kathrin/Sirsch, Jürgen/Wenzelburger, Georg/Zweig, Katharina A. (2022): "Promises and Pitfalls of Algorithm Use by State Authorities", in: Philosophy and Technology, 35:33.

König, Pascal D./Wenzelburger, Georg (2021): "The legitimacy gap of algorithmic decision-making in the public sector: Why it arises and how to address it", in: Technology in Society.

König, Pascal D./Wenzelburger, Georg (2021): „Between technochauvinism and human-centrism: Can algorithms improve decision-making in democratic politics?", in: European Political Science 21:1, 132–149.

Hartmann, Kathrin/Wenzelburger, Georg (2021): „Uncertainty, risk and the use of algorithms in policy decisions: a case study on criminal justice in the USA", in: Policy Sciences 54:2, 269-287.

König, Pascal D./Wenzelburger, Georg (2021): „When Politicization Stops Algorithms in Criminal Justice", in: The British Journal of Criminology, 61:3, 832-851.
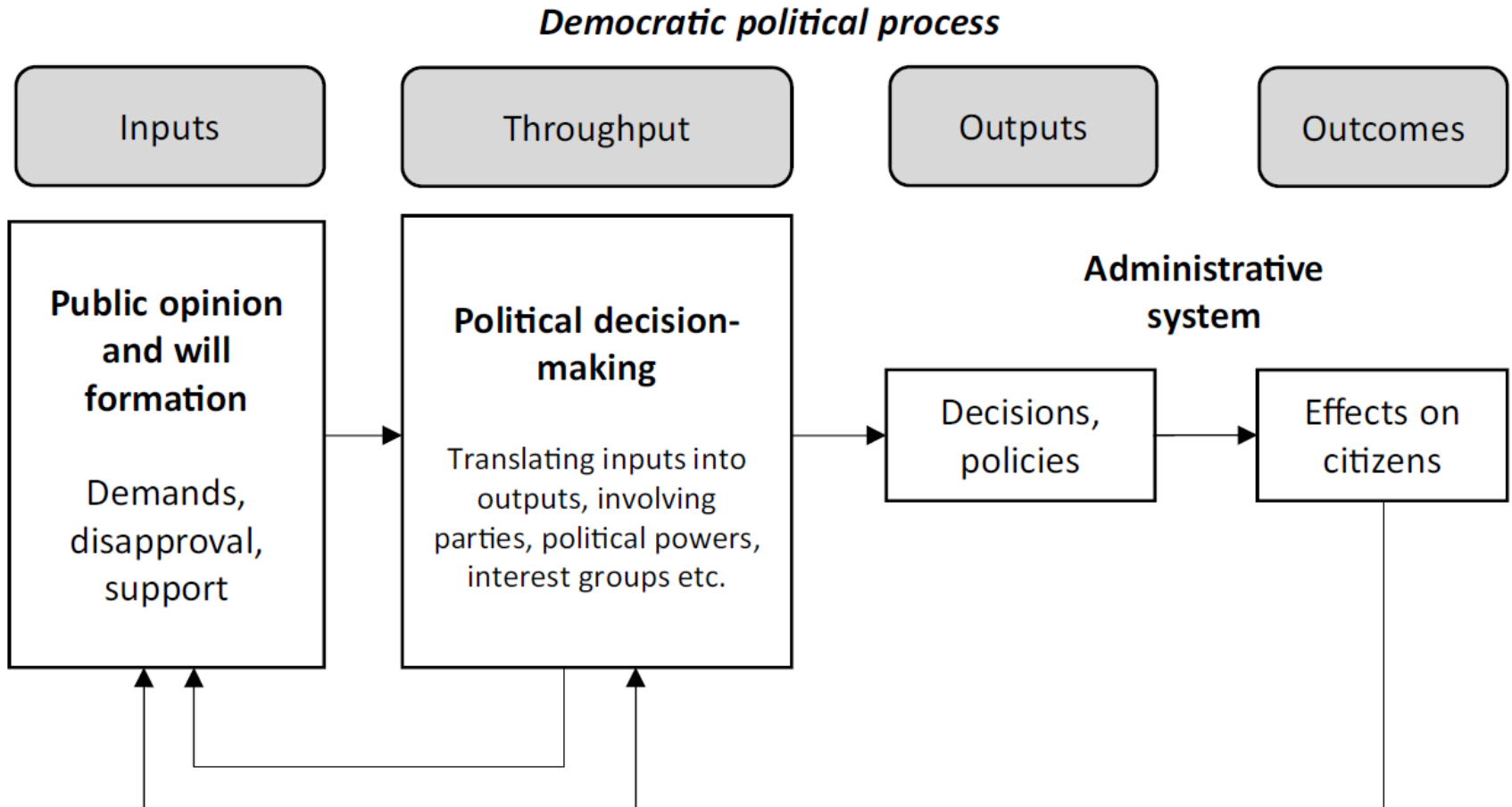
König, Pascal/Wenzelburger, Georg (2020): „Opportunity for renewal or disruptive force? How artificial intelligence alters democratic politics", in: Government Information Quarterly 37:3.

Preliminaries II

- Much of what I will be presenting does not differentiate between the technologies that underlie the algorithm-based tools that are used by public authorities. Hence, when I say AI tools or algorithmic tools, I do not really care about the exact way the tools are made. I am interested in the consequences that the use of algorithmic tools (of whatever nature) has for political and administrative decision-making in democracies.

- However, I do think that there is much to say about data and how it can be used (or not), whether there is unbiased ground truth or to what extent training data from one region can be used in predictions for other regions of the country… → yet, this will not be the focus of this talk!

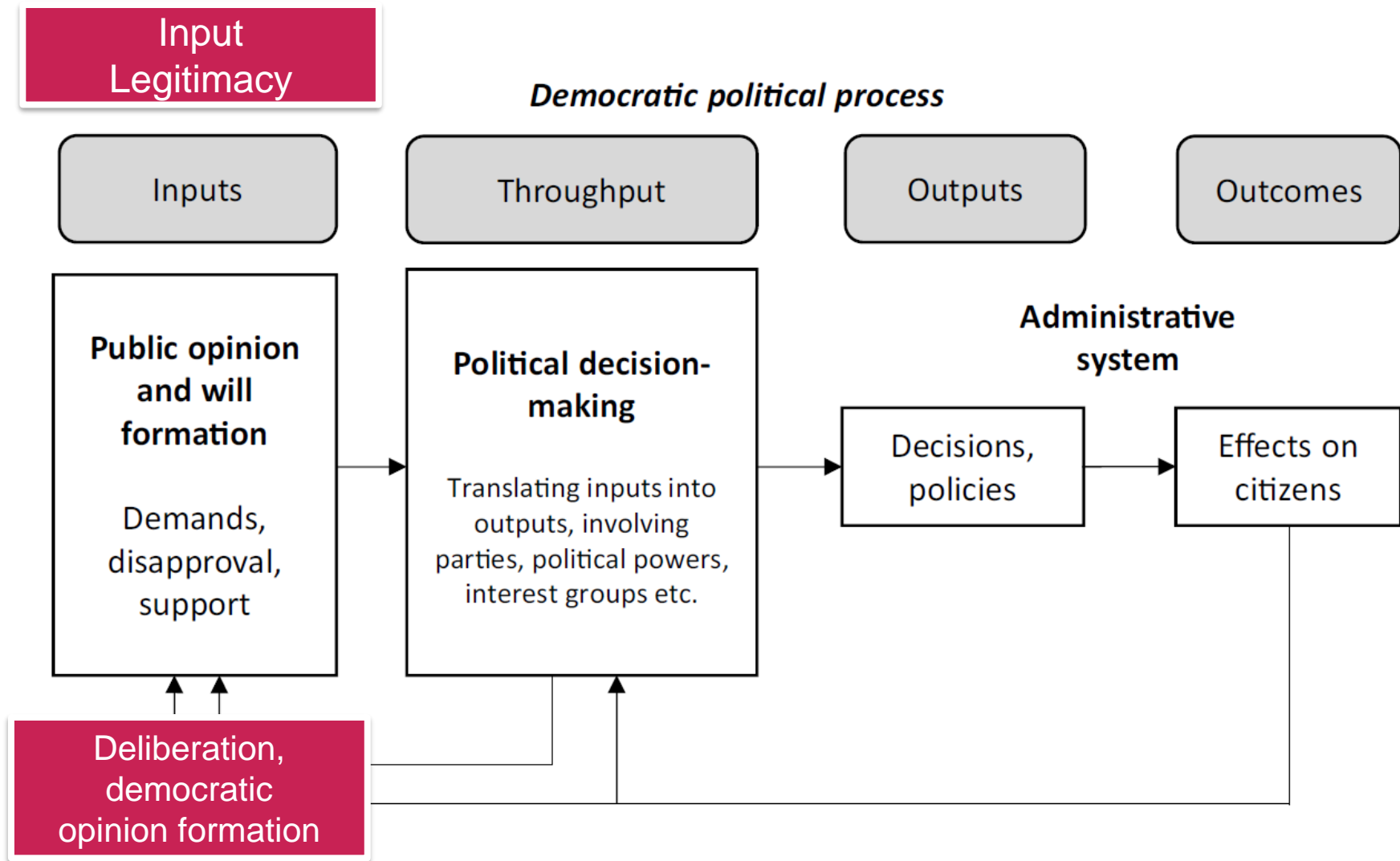# Why should we care about AI and democracy?

A political science perspective
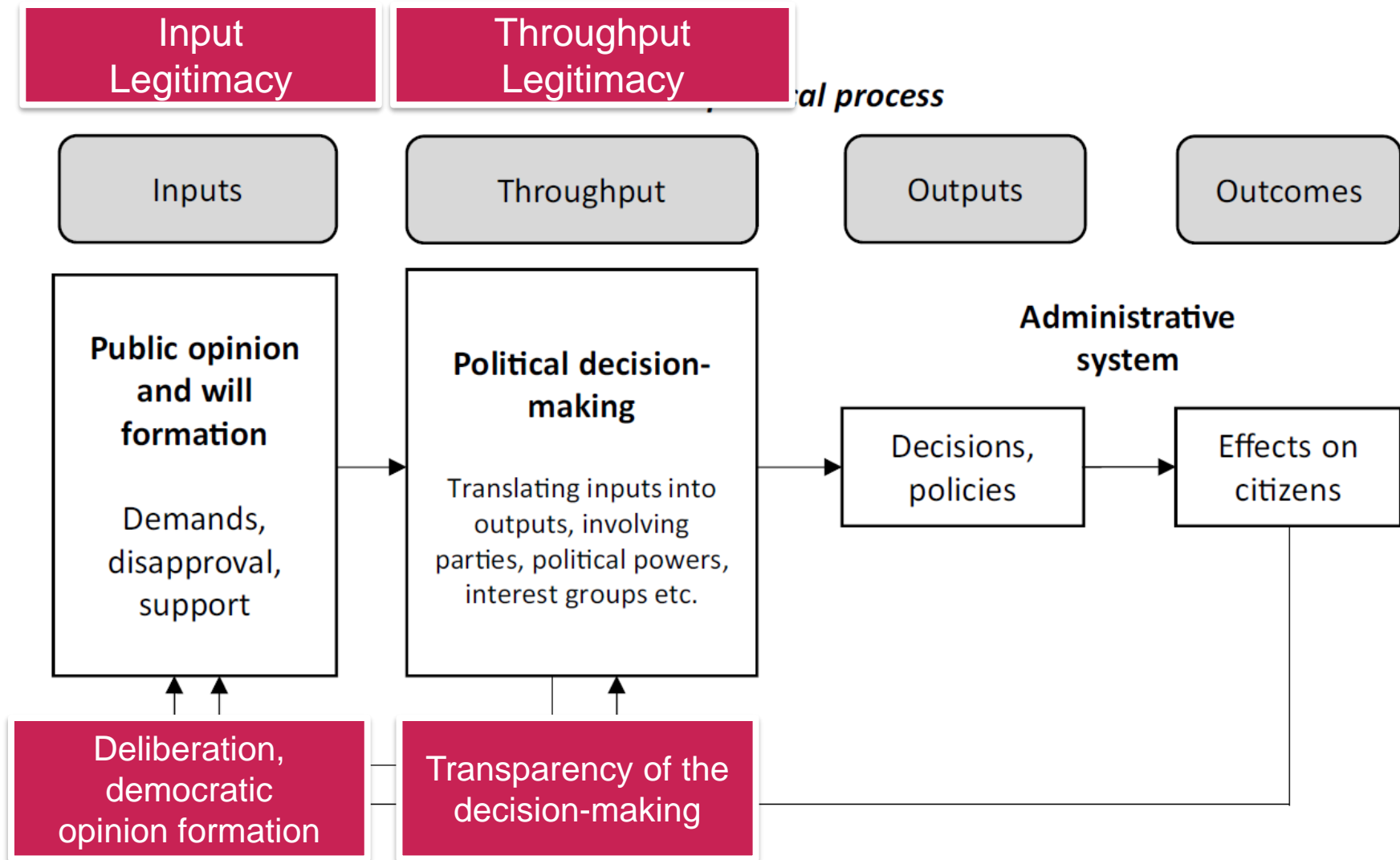
The „Input-Output"-Model of politics

Democratic political process

| Inputs | Throughput | Outputs | Outcomes |

Public opinion and will formation

Demands, disapproval, support

Political decision-making

Translating inputs into outputs, involving parties, political powers, interest groups etc.

Administrative system

Decisions, policies

Effects on citizens

Source: König/Wenzelburger 2020

AI and the Input-Output-Model

Input Legitimacy

Democratic political process

Inputs	Throughput	Outputs	Outcomes

Public opinion and will formation

Demands, disapproval, support

Political decision-making

Translating inputs into outputs, involving parties, political powers, interest groups etc.

Administrative system

Decisions, policies

Effects on citizens

Deliberation, democratic opinion formation

AI and the Input-Output-Model

Input Legitimacy	Throughput Legitimacy

al process

Inputs	Throughput	Outputs	Outcomes

Public opinion and will formation

Demands, disapproval, support

Political decision-making

Translating inputs into outputs, involving parties, political powers, interest groups etc.

Administrative system

Decisions, policies

Effects on citizens

Deliberation, democratic opinion formation	Transparency of the decision-making

Throughput Legitimacy

"Robots take over: Machine to run for MAYOR in Japan pledging 'fair opportunities for all'"

AI and the Input-Output-Model

Input Legitimacy	Throughput Legitimacy	Output Legitimacy

Inputs — Throughput — Outputs — Outcomes

Public opinion and will formation

Demands, disapproval, support

Political decision-making

Translating inputs into outputs, involving parties, political powers, interest groups etc.

Administrative system

Decisions, policies

Effects on citizens

Deliberation, democratic opinion formation

Transparency of the decision-making

Effective problem solving responsive to citizens preferences

Output legitimacy

Denver Pretrial Supervision Guidelines-Misdemeanor and Felony Offenses

Primary Charge and CPAT Category **Enhancers** *(move up one level of supervision)*	•Non VRA Misdemeanors	•NonVRA Felonies	•VRA Misdemeanors	•Misdemeanor DV •Indecent Exposure	•VRA Felony Crimes •Felony DV •DF1 •Burglary of a Dwelling	•Felony Sex Offenses
[→ → →] • Currently supervised on felony probation, parole, or pretrial supervision for any criminal offense • Two or more pending felony cases or misdemeanor assault cases within one year of current offense date • Offense involves a knife or a firearm in current charge • High ODARA score (7+) • Felony Child Abuse or Felony Sex Assault on a Child						
Category 1 *Score: 0 to 17 (87% Success)* *91% Public Safety, 95% Court Appearance*						
Category 2 *Score: 18 to 37 (71% Success)* *80% Public Safety, 85% Court Appearance*						
Category 3 *Score: 38 to 50 (58% Success)* *69% Public Safety, 77% Court Appearance*						
Category 4 *Score: 51 to 82 (33% Success)* *58% Public Safety, 51% Court Appearance*						

KAREN BOGENSCHNEIDER
AND THOMAS J. CORBETT

EVIDENCE-BASED POLICYMAKING

Envisioning a New Era of Theory, Research, and Practice

Second Edition
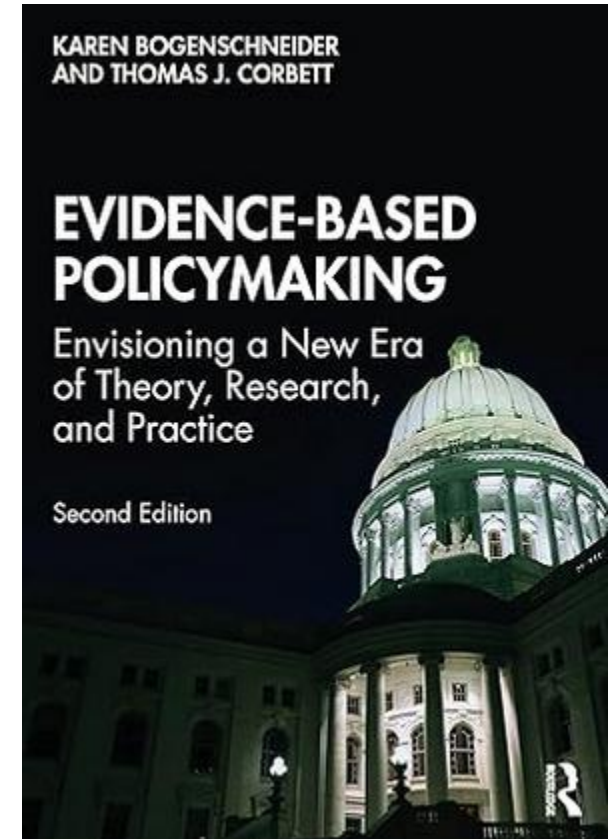
Source: Routledge

AI and the Input-Output-Model

Input Legitimacy | Throughput Legitimacy | Output Legitimacy

Inputs | Throughput | Outputs | Outcomes

Filter bubbles, echo chambers, information diets, micro-targeting and fake news?

Technocratic decisions through „evidence", opaqueness of the process

Transparent and accountable administrative decisions, biased decisions, lack of human discretion?

Deliberation, democratic opinion formation

Transparency of the decision-making

Effective problem solving responsive to citizens preferences

Today's talk: AI use by public authorities
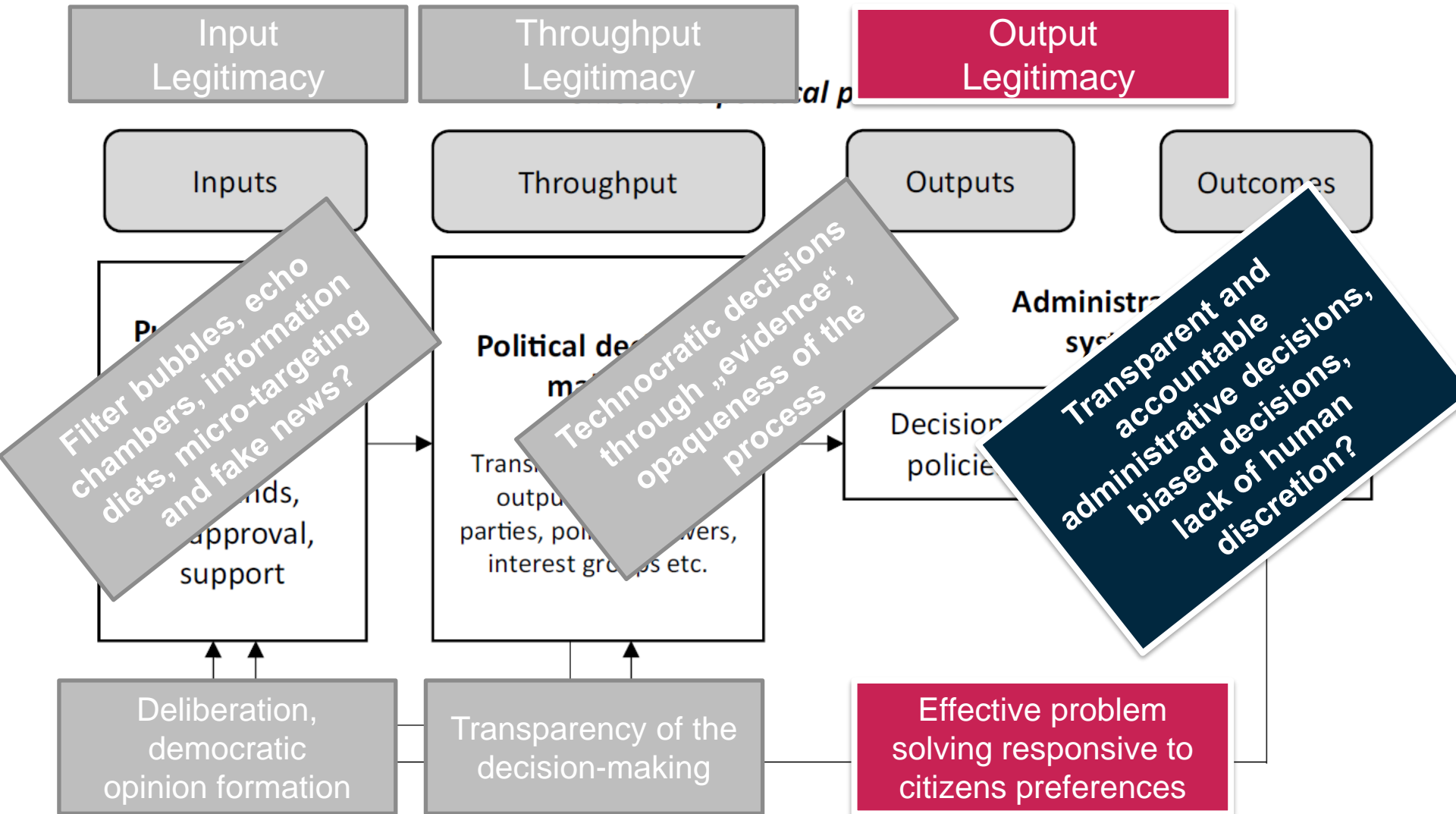
Three main arguments

1) **The use of algorithms in public administration changes the decision-making context of frontline workers.**

2) **Algorithmic tools in public administration are often introduced without questioning the underlying data (and biases)**

3) **Citizens do not care about regulation of algorithmic tools.**

How does AI change
the nature of decision-making
in public administration?

How does AI change frontline work?

1) **Automation bias**: The use of AI changes the decision-making context for frontline bureaucrats as "objective" evidence exists on which decision can be based:

"The process of arriving at a decision changes. The perception of accountability for the final decision changes too. The decision-makers will be inclined to tweak their own estimates of risk to match the model's." (Završnik, 2019: 13)

Završnik, A. (2019). Algorithmic justice: algorithms and big data in criminal justice settings. *European Journal of Criminology*

How does AI change frontline work?

Gigerenzers risk matrix

Nature of unkown	Decision process	Result
Risk (a priori, stochastic) "known unknown"	Deductive (Probabilities)	Objective odds
Risk (statistical)	Inductive (statistical inference, algorithmic predictions)	Correlational patterns, predictions of effects
Uncertainty (fundamental) "unknown unknown"	Heuristic	Intuitive satisficing

Source: adapted from
Mousavi and Gigerenzer 2014, 1673)

How does AI change frontline work?

Gigerenzers risk matrix

Nature of unkown	Decision process	Result
Risk (a priori, stochastic) "known unknown"	Deductive (Probabilities)	Objective odds
Risk (statistical)	Inductive (statistical inference, algorithmic predictions)	Correlational patterns, predictions of effects
Uncertainty (fundamental) "unknown unknown"	Heuristic	Intuitive satisficing

AI based tool

Source: adapted from
Mousavi and Gigerenzer 2014, 1673)

How does AI change frontline work?

1) Automation bias: The use of AI changes the decision-making context for frontline bureaucrats as "objective" evidence exists on which decision can be based:

"The process of arriving at a decision changes. The perception of accountability for the final decision changes too. The decision-makers will be inclined to tweak their own estimates of risk to match the model's." (Završnik, 2019: 13)

2) **Blame Avoidance**: Bureaucrats will tend to follow AI tools, because they may blame them for wrong decisions afterwards, especially in decisions with high risk.

Source: Hartmann/Wenzelburger (2021)

The case: Risk assessment software in CJ

- **Case study on** from an US context with an "extreme" Case (theory building): Risk assessment software (COMPAS) estimating statistical risk based on a machine-learning algorithm

- Method: Exploratory case study, focus on local implementation
  - Interviews with actors on the ground
  - Extensive document analyses

The case: Risk assessment software in CJ

*"And that's to me, what in the end what this is all about: Be as informed as possible **to make as informed a decision as possible**. So at least you can say: **With a clear conscience, you know I utilized the information that is available to me versus: Well I, dart board, I just guess and hope I get it right.**"*

*"I don't like guesses. You know, **you like certainty**. And right, wrong or different. **These validated risk assessment tools bring a level of certainty to the analysis of an individual that we've never had before**"*

Much points in favour of „over-reliance" of bureaucratic decision-makers. While many actors also see the need to be well informed, most emphasize the easy-to-implement tool that helps decisions.

Not much evidence that blame avoidance matters, but other results from experiments (Feier et al. 2022).

Feier, T., Gogoll, J. & Uhl, M. Hiding Behind Machines: Artificial Agents May Help to Evade Punishment. *Sci Eng Ethics* **28**, 19.

Three main arguments

1) **The use of algorithms in public administration changes the decision-making context of frontline workers.**

   ➔ Frontline workers are often happy to use algorithmic tools (caveat: algorithm aversion), because they see them as helpful instruments to increase the certainty of their decision-making.

Three main arguments

1) The use of algorithms in public administration changes the decision-making context of frontline workers.

2) **Algorithmic tools in public administration are often introduced without questioning the underlying data (and biases)**

3) **Citizens do not care about regulation of algorithmic tools.**

Algorithmic tools in public administration are often introduced without questioning the underlying data (and biases)

The evidence-based policy dream

- **Starting from the NPM-wave** in the 1980s/1990s (rooted in neo-liberal economic policies), governments have sought to make administration more efficient.

- The use of algorithmic tools has been seen as a way to take decisions in public administration more efficiently.

- Biases have been downplayed, as human decision-making is also known to be biased.

The evidence-based policy dream

- **Case study on the introduction of "AMAS"** as a tool to inform decision-making in Austrian labor market agencies

- Predicts three classes of "employability" of job seekers depending on their "history" (and, relatedly, socio-demographics).

- Job seekers with low chances of re-integration are no longer offered re-training measures for labor market integration

- Government pursued the reform for cost saving arguments against push-back from trade-unions or NGOs
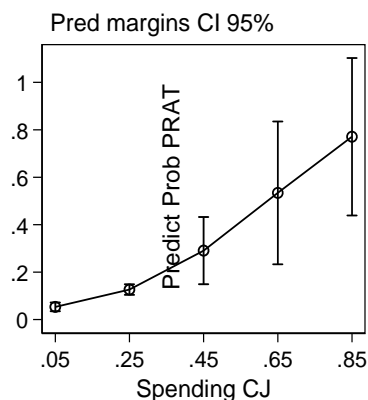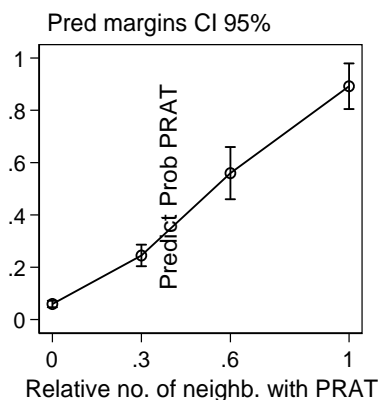
The evidence-based policy dream

- **Comparative study** on the introduction of pre-trial risk assessment tools in US states.
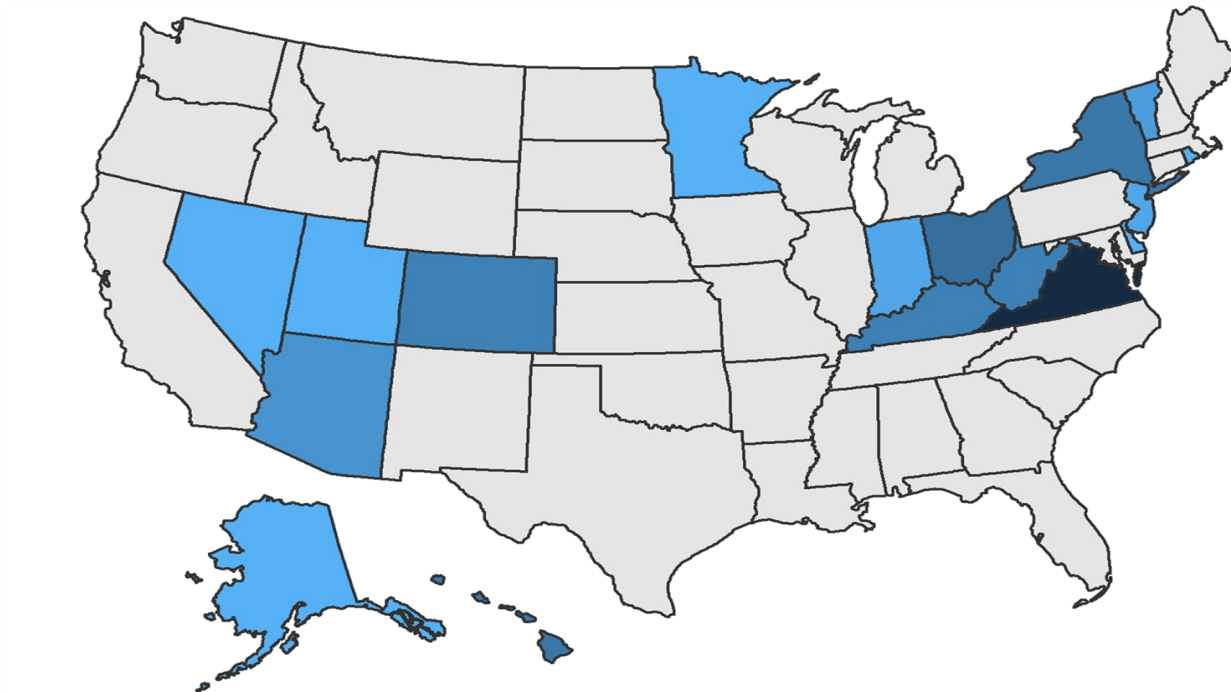
Source: König/Wenzelburger 2022

The evidence-based policy dream

- **Main reasons for adopting the tools: Mainly money…**
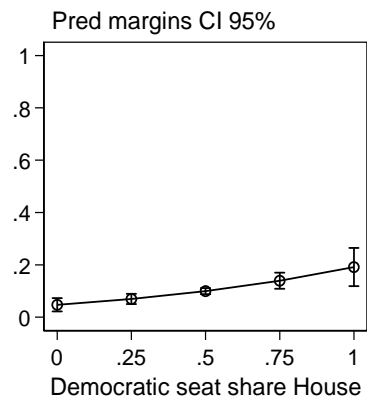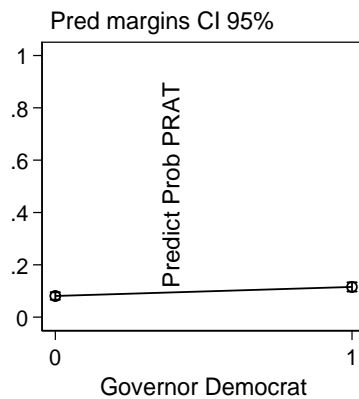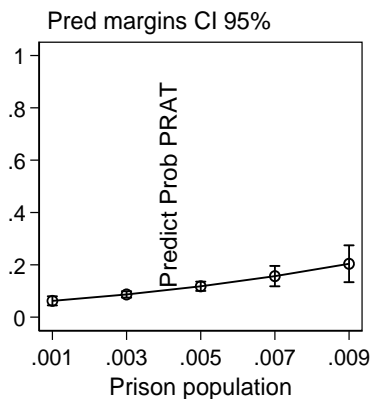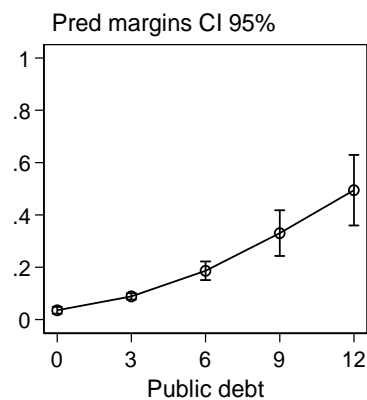
The evidence-based policy dream

- **Case study on the introduction of the Virginia pretrial tool:** Main reason for introducing the tool was the hope to keep the number of prisoners under control, as a reform of parole was estimated to increase the number or prisoners.

"I do think both sides at the time recognized they— we needed to avoid prison overcrowding. And so, I think that was part of the willingness to figure out a way to create that balance, because otherwise we would have a mess that Florida had already created."

Source: Wenzelburger 2025

The evidence-based policy dream

- **Questions of bias or false predictions are hardly discussed – unless they become politicized:**

  - Reform of an automated university admission system (not AI, but matching algorithm!) in France "APB" was abolished when it was clear that "random draws" were sometimes used to match students to university programs.

  - Algorithmic tools already in use in three US states were abolished because of
    - An increase in violent crime (framed as result of the tool letting too many "risky" people go free) (Alaska, Iowa)
    - Cases of strong racial bias considered unfair (Idaho)

Source: Wenzelburger 2021

Three main arguments

2) **Algorithmic tools in public administration are often introduced without questioning the underlying data (and biases)**

➜ Political decisions on the adoption of AI tools are not driven by technical features or the quality of prediction, but by political dynamics, e.g. preferences for lower public spending in CJ (US) or being seen as "modernizers" (Austria)

Three main arguments

1) The use of algorithms in public administration changes the decision-making context of frontline workers.

2) Algorithmic tools in public administration are often introduced without questioning the underlying data (and biases)

3) **Citizens do not care about regulation of algorithmic tools.**

Citizens do not care about regulation of algorithmic tools.
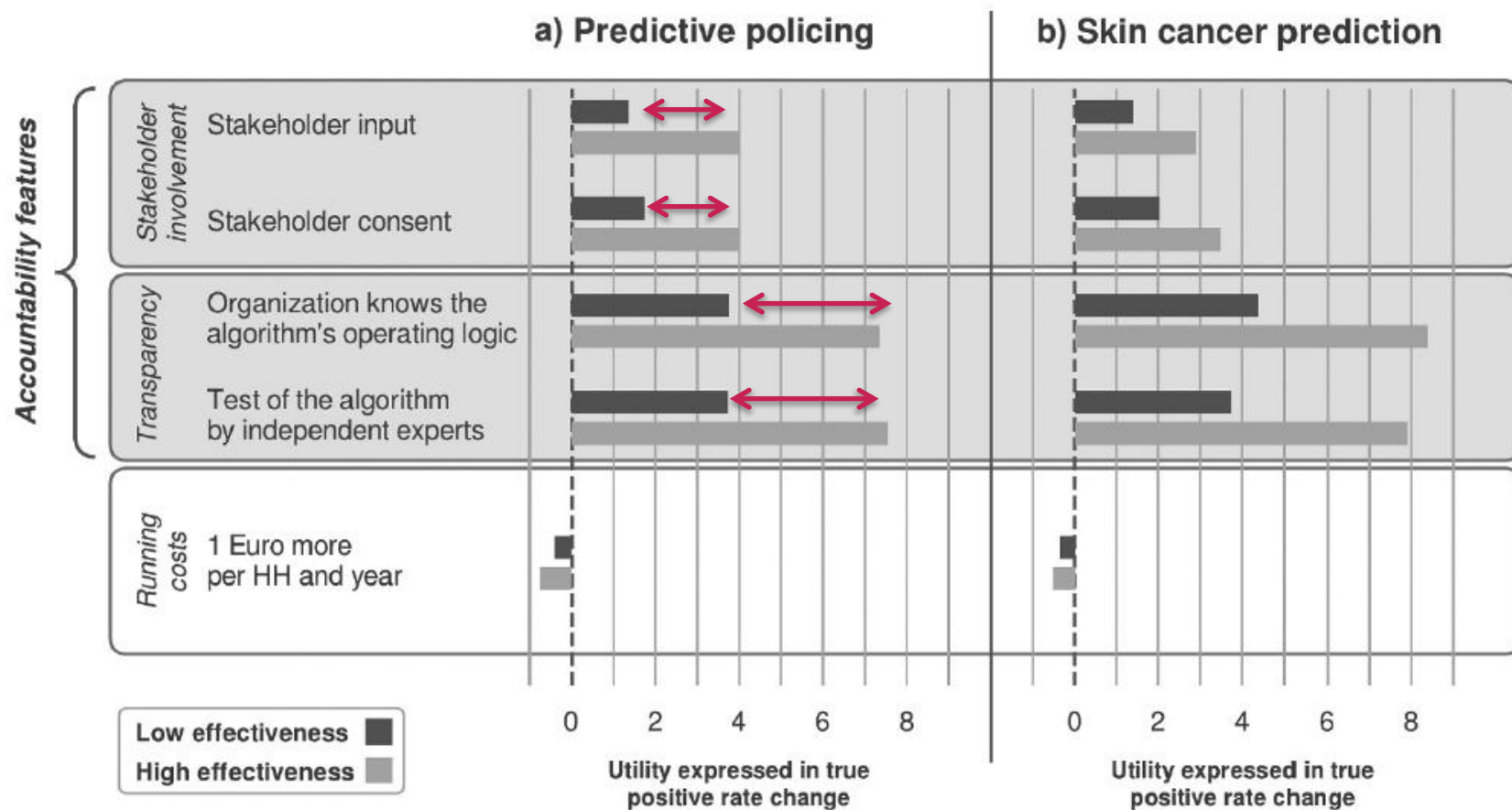
Citizens and views on AI use by the state

- Survey experiment asking under what conditions respondents are ready to trade off more regulation of AI (e.g. stakeholder involvement, transparency, expert oversight, etc.) against efficiency gains.

- **Reference:** No stakeholder involvement, no transparency, effectiveness (true positive rate) of 5% recall (!) and running costs of 6 Euros per household per year.

König et al. (2024); Wenzelburger et al. (2024)

Citizens and views on AI use by the state
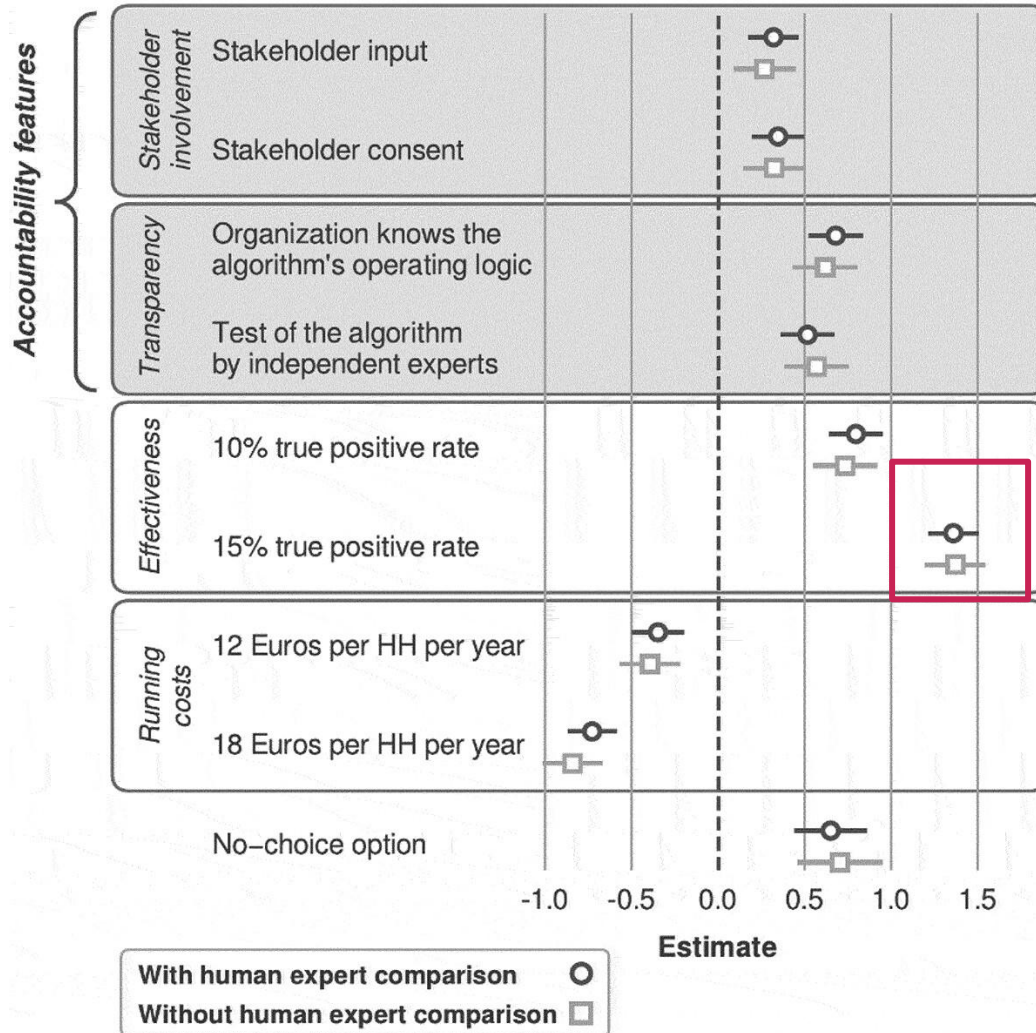
Strongest predictor of preferences for an AI use in predictive policing is its **effectiveness**!

König et al. (2024); Wenzelburger et al. (2024)

Citizens and views on AI use by the state

König et al. (2024); Wenzelburger et al. (2024)

Citizens and views on AI use by the state

... And the **salience of the policy area** for the respondent also increases the support for the use of an algorithm (independently from effectiveness!)

TABLE 2 Coefficient estimates from OLS regression predicting support for algorithm use

	Predictive policing		Skin cancer prediction	
	Model 1	Model 2	Model 3	Model 4
Main predictors				
Personal importance of addressed problem (H1A)	0.17**	0.19**	0.17**	0.18**
	(4.35)	(4.55)	(3.80)	(3.63)
Worried burglary/skin cancer (H1B)				
Not at all	Reference category		Reference category	
A bit worried	0.04*		0.04*	
	(2.35)		(2.32)	
Quite worried	0.04		0.07**	
	(1.91)		(3.32)	
Very worried	0.05		0.09**	
	(1.72)		(2.86)	

König et al. (2024); Wenzelburger et al. (2024)

Three main arguments

3) Citizens do not care about regulation of algorithmic tools.

➔ Citizens do prefer transparency and accountability mechanism, but they are clearly less important to them than effectiveness and relevance of the policy area. Overall, this points to citizens that mainly prefer the promise of efficient solutions to problems they care about and it does not really matter much how they are achieved.

Conclusion

Conclusion

- Algorithms (based on AI or not) are increasingly used by state authorities and support administrative decision-making

- Bureaucrats are (mostly) happy to work with them as they seem to provide evidence to come to less uncertain decisions (e.g. parole decisions, training of unemployed)

- The political decision to use algorithms is not mainly based on considerations of performance or bias, but driven by political dynamics

- It cannot be expected from citizens to push for regulation or transparency

- However, from a point of view of **democracy**, we would need at least **transparent and accountable** decisions of public administration. Biased decisions will always happen but should not be in-built systematically and protected by commercial secrecy...
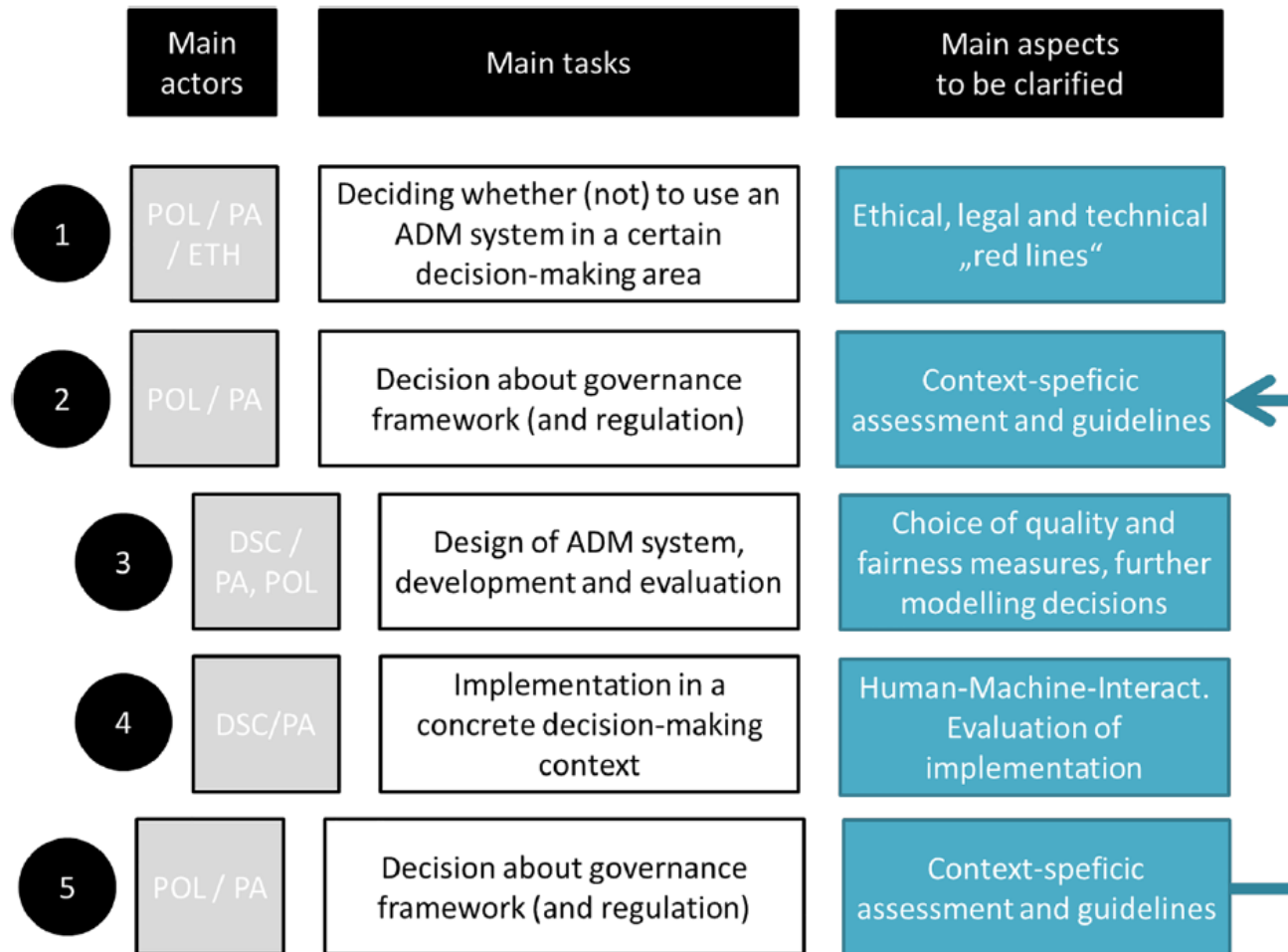
A step further: Interdisciplinary teams

Fig. 1 Stages in a decision process leading to introduction of an ADM. ETH ethics; DSC data science; PA public administration; POL politics

Source: Haeri et al. 2022

Many thanks for listening!