

Responsible use of tools

Trust in science

Shaping the Digital Future of ErUM Research: Sustainability & Ethics

31.07.2025

Jan M. Bürger



Dataedo /cartoon

Proh@Dataedo



ar.inspiredpencil.com



Main important points for a tool

1. It works correct (and efficient)
2. It fits to the problem
3. No bias in the train data
4. I use it correct

But: What can I do to proof the points?

→ **A tool fits to the problem and has no bias – but how to proof?**



The example of this talk

- No physical model
- We use: The Wahl-O-Mat of the Bundestagswahl 2021/25
 - How does it work?
 - How can I analyze it?
 - Does it work?
- Responsibility for society

→ **Have a look into the Wahl-O-Mat of the Bundestagswahl 2021/25**



The Wahl-O-Mat

- Roughly speaking: election recommendation
- Goal: Give an orientation, which party fits best to the user
- The Wahl-O-Mat provides statements and calculate the matches between parties and the user



→ **Wahl-O-Mat compare answers of statements between the user and different parties**



Important: Good choice of the statements

- Assume larger parties have more agreement with majority
- Not only statements, where one party fits the majority
 - Boring, this party will be the best match for most of user

→ **A good selection of statements is the crucial point**



How to proof?

- Use a LLM (here Microsoft Copilot)
 - Direct way: “Compare for each statement the party XY with the majority meaning.”
- Calculate with Excel
- Only parties of Bundestag

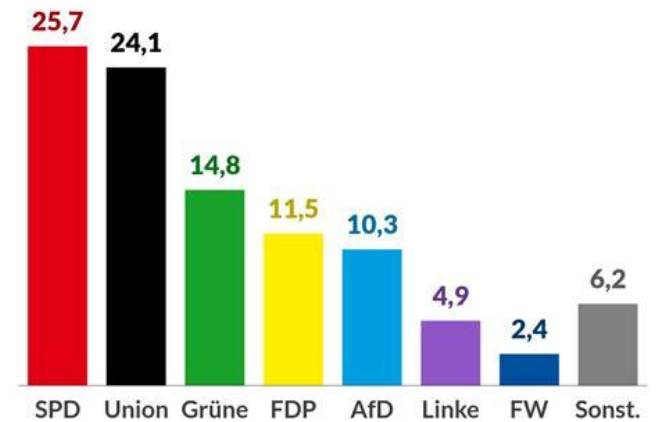
→ **Ask LLM to compare parties with the majority (based on statements)**



Wahl-O-Mat 2021: Close to be good

- # fitting statements nearly equal
- *Some gaussian*
 - (χ^2 : 0,48; mean 19, std 1)
- Larger parties less statements with high consistency with majority

Bundestagswahl 2021: vorläufiges Ergebnis



Quelle: Der Bundeswahlleiter, Stand: 27.09.2021
Diagramm: www.bundestagswahl-2021.de

Method 1 2021	BSW	AFD	CDU	FDP	SPD	Grüne	Linke
High Consistency	N.N.	16	19	18	19	20	21
Low Consistency	N.N.	22	19	20	19	18	17

Table: # Statements (of 38) a party fits the majority meaning

→ Choose of statements seems good (not perfect)



Wahl-O-Mat 2025: Different to election

- 3 parties with high number of statements fitting the majority
- *No gaussian*
- Larger parties less statements with high consistency with majority



Method 1 2025	BSW	AFD	CDU	FDP	SPD	Grüne	Linke
High Consistency	N.N.	17	19	20	28	30	30
Low Consistency	N.N.	21	19	18	10	8	8

Table: # statements (of 38) a party fits the majority meaning

→ **No correlation between statements of the Wahl-O-Mat and the election**



Conclusion

- Maybe some bias in statements (2025 more than 2021)
- But also the analytic can be biased
 - (one) LLM
 - Prompt
 - ..

→ Interest in discussion, more research needed

- And now?
 - Have a responsible look on tools, prof them
 - Be a good example, reinforce the trust in science

→ **Be careful with tools: You have a responsibility – also for society**





Possible problems: Wahl-O-Mat

- Decision of 38 (of 80) after answer of parties
- Trust in answers of parties (instant of parliaments sessions)
- Leading questions

→ **Choose of statements can be the problem**



Possible problems: Analytic

- LLM as database
- Only one LLM
- LLM prompt
 - „Compare for each statement the party XY with the majority meaning.”
 - Nearby
 - Many response
 - Trust in database of LLM about parties
 - „Meaning of majority of statements”, database of Wahl-O-Mat for parties
 - Only on LLM response
 - Compare every party with same database (more equal)
 - Need to trust the parties answers
- Assumption of distribution
 - **Analytic with LLM can be also biased**



Possibility for a problem

- Assume: Some positions of a party fits/fits not to the majority
- Choosing mostly statements, where one party fits to majority:
This party get best results for most of the users

1: agree, 0: neutral, -1: not agree

	Positions			Distance		Points (2-Distance)	
These	User	Party A	Party B	Party A	Party B	Party A	Party B
These A	1	1	0	0	1	2	1
These B	-1	0	-1	1	0	1	2
These C	0	1	1	1	1	1	1

#

Questions Max. Points

3	6These A-C	0,667	0,667
2	4These B-C	0,5	0,75
1	2These B	0,5	1

- If the statements are biased, the results are not helpful
- Choose statements carefully



Assumptions

- All parties have positions, where they fitting the majority
- For success: Not only „fitting“ statements of one party
 - The result is clear: Most matches with this party → not useful

→ Assume following

- All parties ideally equal number of „matching statements“
- Larger parties allow more „matching statements“
- Allow some gaussian (\sim std 1, mean 19 Questions=38/2)

→ **Use nearly equal “matching statements” for all parties, larger maybe some more**



Methods

- Briefly analytic based on Microsoft Copilot
- Calculation with excel
- Only parties of Bundestag
- 2 methods, to crosscheck the methods

→ **Ask Microsoft Copilot for agreement between majority and parties**



Compare both methods

Directly compare parties with majority

- Heavy to calculate (AI response for each party)
 - Not really sustainability
- Not longer supported from Microsoft Copilot (so no data for BSW)

Copilot sagte

Hier ist eine Übersicht der Thesen aus dem Wahl-O-Mat zur Bundestagswahl 2025 und wie die Positionen der CDU mit den Ergebnissen aus Bevölkerungsumfragen übereinstimmen:

1. **Ausbau der erneuerbaren Energien:** Hohe Übereinstimmung
2. **Einführung einer CO2-Steuer:** Geringe Übereinstimmung
3. **Beibehaltung des Atomausstiegs:** Geringe Übereinstimmung
4. **Verbot von Einwegplastik:** Hohe Übereinstimmung
5. **Abschaffung des Solidaritätszuschlags:** Hohe Übereinstimmung
6. **Einführung eines bedingungslosen Grundeinkommens:** Geringe Übereinstimmung
7. **Abschaffung des Mindestlohns:** Hohe Übereinstimmung

Ask only to the agreement of the majority to the statements

- Wahl-O-Mat offers answers of each party
- Only one AI response
- All parties have the same database (not bias by response)
 - Have to trust in answers of the parties
- More flexible (e.g. adding parties)

Wahl-O-Mat®

Bundestagswahl 2021

Vergleich der Positionen

	CDU CSU	SPD	AfD
1. Auf allen Autobahnen soll ein generelles Tempolimit gelten.	✗	✓	✗
2. Deutschland soll seine Verteidigungsausgaben erhöhen.	✓	✓	✓
3. Bei Bundestagswahlen sollen auch Jugendliche ab 16 Jahren wählen dürfen.	✗	✓	✗
4. Die Förderung von Windenergie soll beendet werden.	✗	✗	✓
5. Die Möglichkeiten der Vermieterinnen und Vermieter, Wohnungsmieten zu erhöhen, sollen gesetzlich stärker begrenzt werden.	○	✓	✗
6. Impfstoffe gegen Covid-19 sollen weiterhin durch Patente geschützt sein.	✓	✓	✓
7. Der für das Jahr 2038 geplante Ausstieg aus der Kohleverstromung soll vorgezogen werden.	✗	✓	✗

→ Method 1 is more direct, Method 2 probably better/more efficient
→ But: Both should be give similar results



Results: Compare methods

It seems to very not equal distributed

Method 1 2021	BSW	AFD	CDU	FDP	SPD	Grüne	Linke
High Consistency	N.N.	16	19	18	19	20	21
Low Consistency	N.N.	22	19	20	19	18	17

χ^2 : 0,48
(Gauß, Mean 19, std 1)

Method 1 2025	BSW	AFD	CDU	FDP	SPD	Grüne	Linke
High Consistency	N.N.	17	19	20	28	30	30
Low Consistency	N.N.	21	19	18	10	8	8

χ^2 : 0
(Gauß, Mean 19, std 1)

Method 2 2025	BSW	AFD	CDU	FDP	SPD	Grüne	Linke
# identic	20	7	14	11	24	26	22
# different	18	31	24	27	14	12	16

Method 2 2025 numeric	BSW	AFD	CDU	FDP	SPD	Grüne	Linke
Distance 0	20	7	14	11	24	26	22
Distance 1	12	9	9	10	9	7	10
Distance 2	6	22	15	17	5	5	6
SumPRODUKT	24	53	39	44	19	17	22

→ Both methods are correlated, but not equal

→ Results are not equal for each party: Some parties higher agreement with majority



Finally: Connected to size of parties?

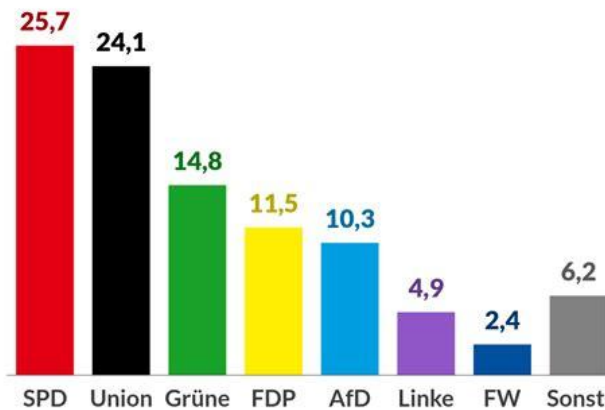
Does it fits with the results of the Bundestagswahl?

Method 1 2021	BSW	AFD	CDU	FDP	SPD	Grüne	Linke
High Consistency	N.N.	16	19	18	19	20	21
Low Consistency	N.N.	22	19	20	19	18	17

Method 1 2025	BSW	AFD	CDU	FDP	SPD	Grüne	Linke
High Consistency	N.N.	17	19	20	28	30	30
Low Consistency	N.N.	21	19	18	10	8	8

Method 2 2025 numeric	BSW	AFD	CDU	FDP	SPD	Grüne	Linke
Distance 0	20	7	14	11	24	26	22
Distance 1	12	9	9	10	9	7	10
Distance 2	6	22	15	17	5	5	6
SumNPRODUKT	24	53	39	44	19	17	22

Bundestagswahl 2021: vorläufiges Ergebnis



Quelle: Der Bundeswahlleiter. Stand: 27.09.2021
Diagramm: www.bundestagswahl-2021.de



Quelle: Bundeswahlleiterin

- 2021: Wahl-O-Mat and Bundestagswahl not as strong connected
- 2025: big difference – seems other way as expected



Conclusion

Expectation	2021	2025
Agreement with majority for each party equal (let say half of 38 questions)	Close (s. gaussian)	Absolutely not
Random: Allow gaussian (std 1, mean 19)	Good enough (one party one to small)	Fare away
Larger parties allow some higher agreement	Not really (but not absolutely clear)	Seems more other way around

- Maybe, some bias in statements (especially in 2025)
- Related questions:
 - Was there a impact of the Bundestagswahl?
 - How to improve?
 - Leading questions (more yes/no questions of a party)
- And now?
 - Have a responsible look on tools, prof them
 - Be a good example, reinforce the trust in science

→ **Be careful with tools: You have a responsibility – also for society**





Sidenote: Leading questions

- Leading questions have maybe a impact (More „anti“ or „pro“)
- Not focus of this talk

	Bevölkerung	BSW	AFD	CDU	FDP	SPD	Grüne	Linke
# Stimme zu	22	16	16	19	14	20	20	18
# neutral	8	8	5	3	4	3	1	2
# stimme nicht zu	8	14	17	16	20	15	17	18
Sum	38	38	38	38	38	38	38	38

→ It is not equal – maybe a impact of the results



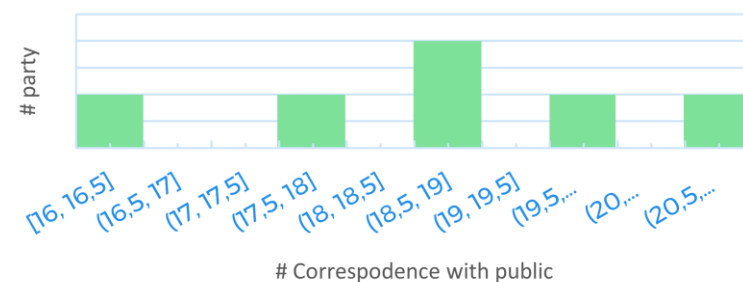
Focus on this talk: Wahl-O-Mat

- Introduction to Wahl-O-Mat
- Expectations
- Methods
- Results

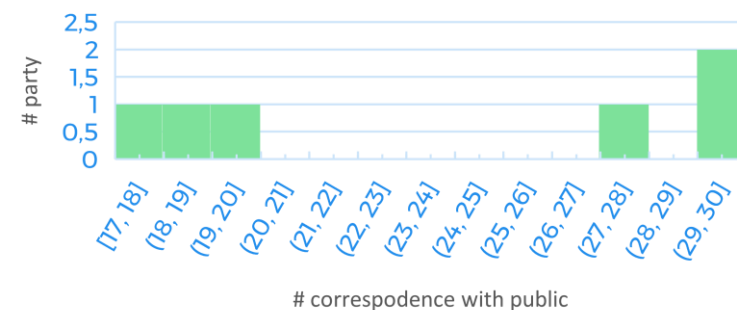
Method 1 2021	AFD	CDU	FDP	SPD	Grüne	Linke
High Consistency	16	19	18	19	20	21
Low Consistency	22	19	20	19	18	17
Sum	38	38	38	38	38	38

Method 1 2025	AFD	CDU	FDP	SPD	Grüne	Linke
High Consistency	17	19	20	28	30	30
Low Consistency	21	19	18	10	8	8
Sum	38	38	38	38	38	38

Wahl-O-Mat 2021

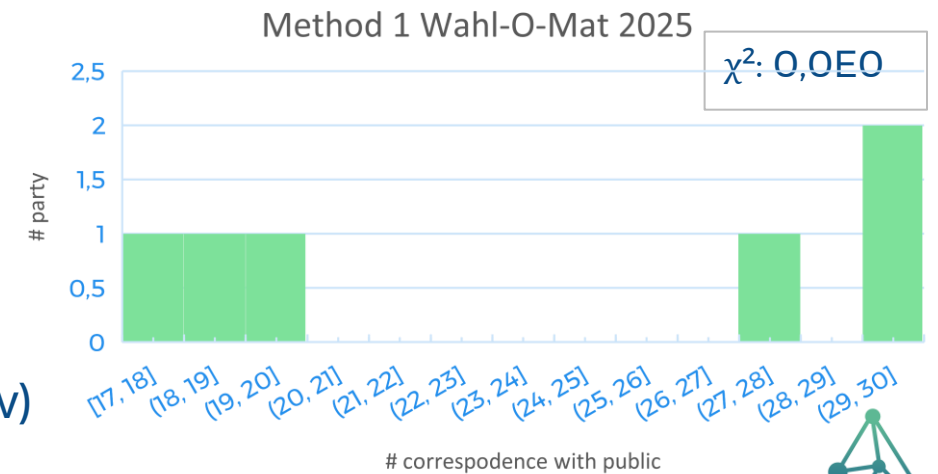
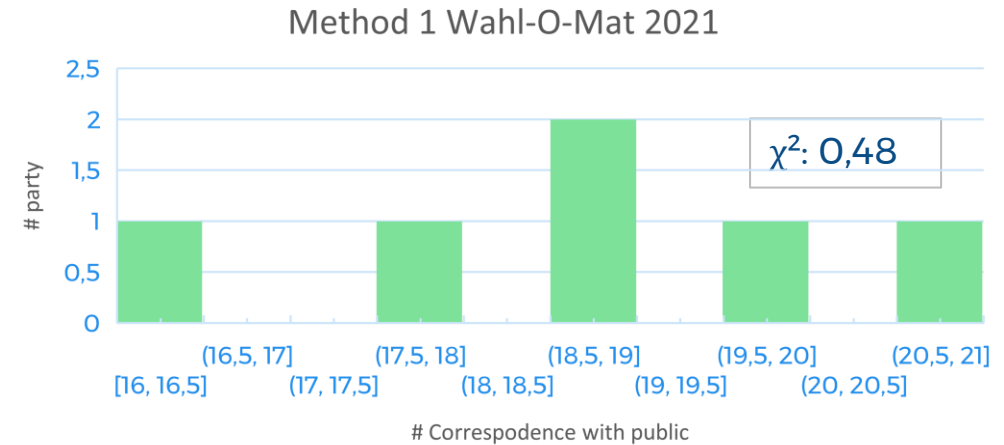


Wahl-O-Mat 2025



Method 1: Is it Gaussian? Compare 2021/25

- Remember: Example Wahl-O-Mat – expectations
 - Objective/useful: Correspondence of positions of the party with a large part of the population nearly equal (half of the questions (38/2))
 - Random: larger parts higher agreement, (or about Gaussian curve with std 1 around 19)



→ 2021: looks good (just the lowest party is 1 question to low)
→ 2025: interesting (especially: x-axis starts at 17..)



Results: Deeper look in each parties

It seems to very not equal distributed

Method 1 2021	AFD	CDU	FDP	SPD	Grüne	Linke
High Consistency	16	19	18	19	20	21
Low Consistency	22	19	20	19	18	17
Sum	38	38	38	38	38	38

Method 1 2025	AFD	CDU	FDP	SPD	Grüne	Linke
High Consistency	17	19	20	28	30	30
Low Consistency	21	19	18	10	8	8
Sum	38	38	38	38	38	38

Method 2 2025	BSW	AFD	CDU	FDP	SPD	Grüne	Linke
# identic	20	7	14	11	24	26	22
# unterschiedlich	18	31	24	27	14	12	16
Sum	38	38	38	38	38	38	38

Method 2 2025 numeric	BSW	AFD	CDU	FDP	SPD	Grüne	Linke
Distance 0	20	7	14	11	24	26	22
Distance 1	12	9	9	10	9	7	10
Distance 2	6	22	15	17	5	5	6
Sum	38	38	38	38	38	38	38
SUMENPRODUKT	24	53	39	44	19	17	22

→ Both methods are correlated, but not equal

→ All analysis indicated, that the Wahl-O-Mat seems to be not equal

