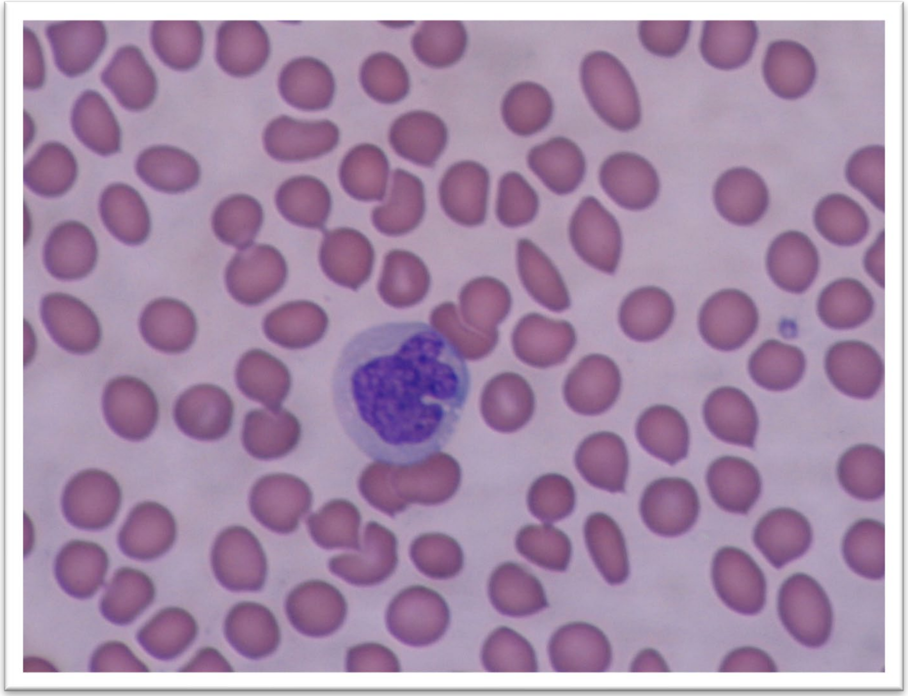
Rethinking Usability Heuristics for Modern Biomedical Interfaces

Stefan Röhrl, M.Sc.
DigitalWorld 2023 Congress
16th International Conference on Advances in Computer-Human Interactions

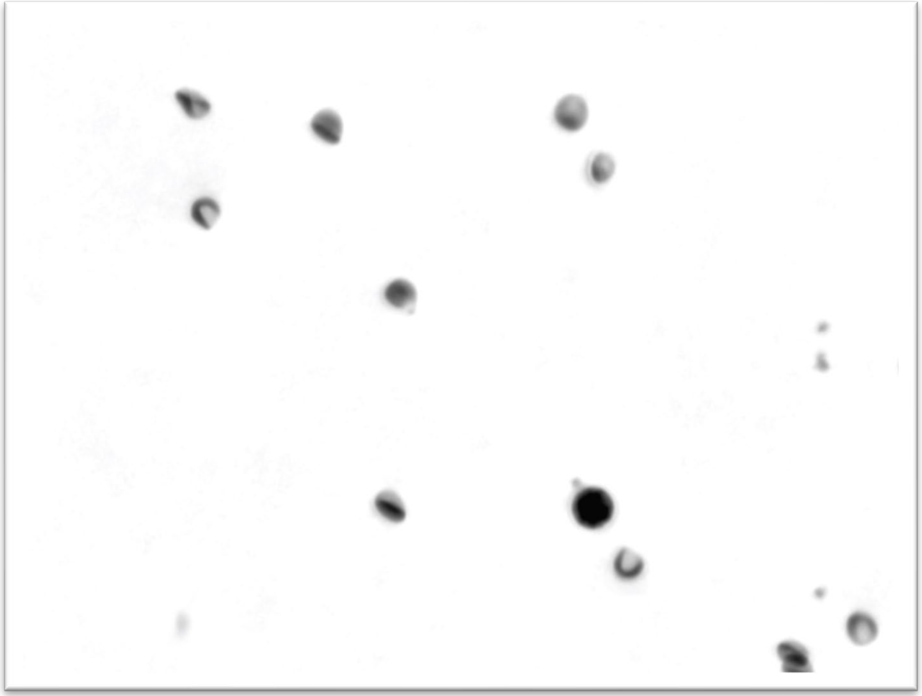
Holographic Cell Imaging

Stained Light Microscope Image

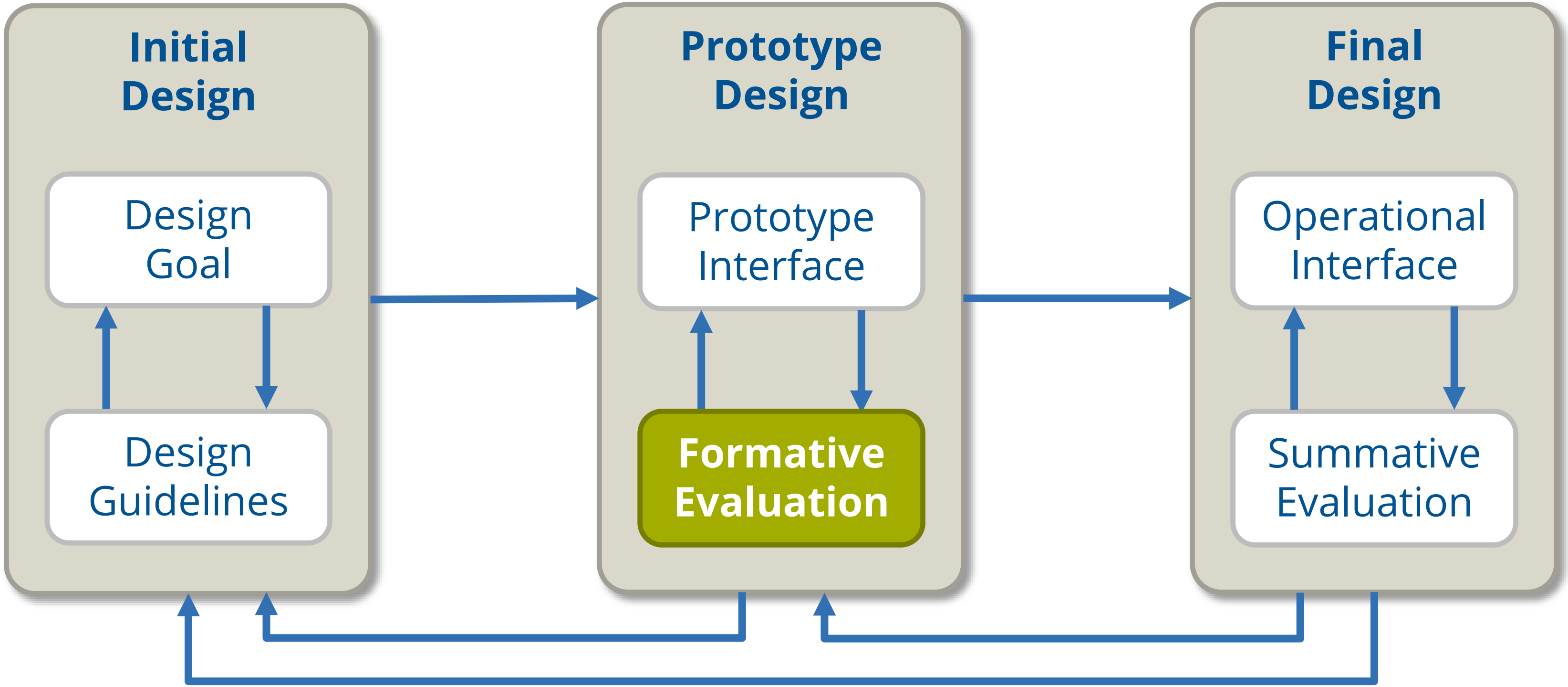


https://en.wikipedia.org/wiki/Agranulocyte#/media/File:Monocyte_no_vacuoles.JPG

Holographic Phase Image

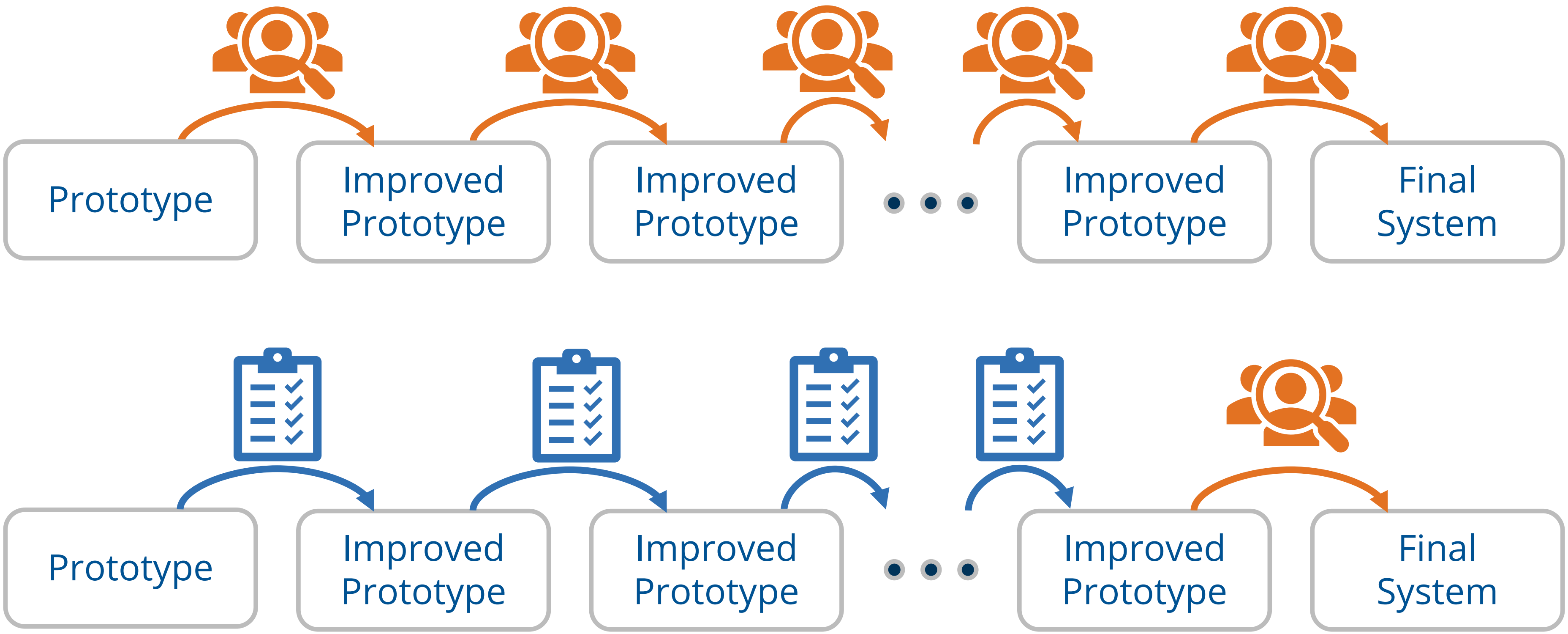


User Interface Development Process



Kies (1998),
Hartson (2001)

Speeding up Formative Evaluation



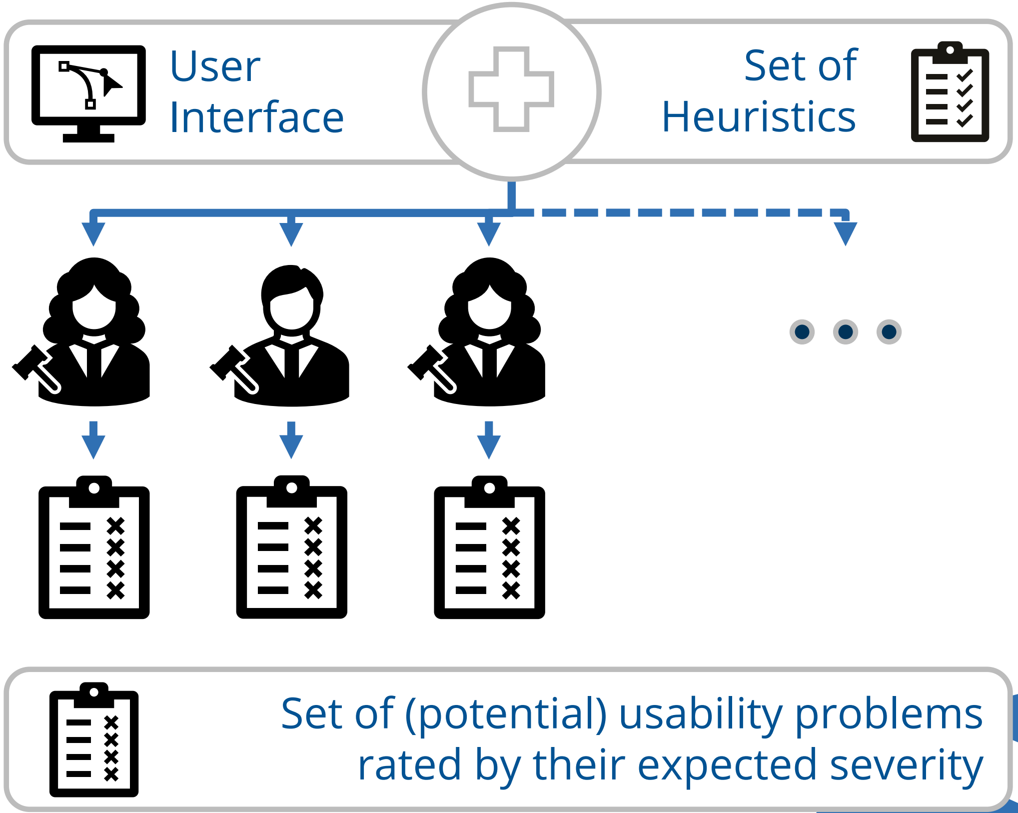
Heuristic Evaluation by Experts

Method *Nielsen (1990,1995)*

- Predefined best practice rules
- Team of (five) expert evaluators
- Rate severity of violations [0, 4]
- Overall Summary

Advantages *Nielsen (1993), Noyes (1999)*

- Time-efficient
- Less Human Resources
- Can be performed by novices and experts equally



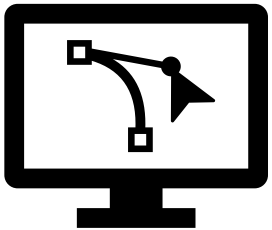
Goals

Custom Heuristic Rules

- Develop a set of custom heuristics
- Focus on biomedical user group
- Compare their performance against other similar and popular heuristics

User Interface

- Support interdisciplinary research
- Allow interaction of humans and AI
- Fully functioning prototype



Existing Heuristics

Nielsen's heuristics *Nielsen (1990,2005)*

1	Simple and natural dialogue
2	Speak the user's language
3	Minimize the user's memory load
4	Consistency and standards
5	Feedback
6	User control and freedom
7	Shortcuts
8	Good error messages
9	Error prevention
10	Help and documentation

Human-AI interaction heuristics *Amershi (2019)*

1	Make clear what the system can do
2	Make clear how well the system can do what it can do
3	Time services based on context
4	Show contextually relevant information
5	Match relevant social norms
6	Mitigate social biases
7	Support efficient invocation
8	Support efficient dismissal
9	Support efficient correction
10	Scope services when in doubt
11	Make clear why the system did what it did
12	Remember recent interactions
13	Learn from user behavior
14	Update and adapt cautiously
15	Encourage granular feedback
16	Convey the consequences of user actions
17	Provide global controls
18	Notify users about changes

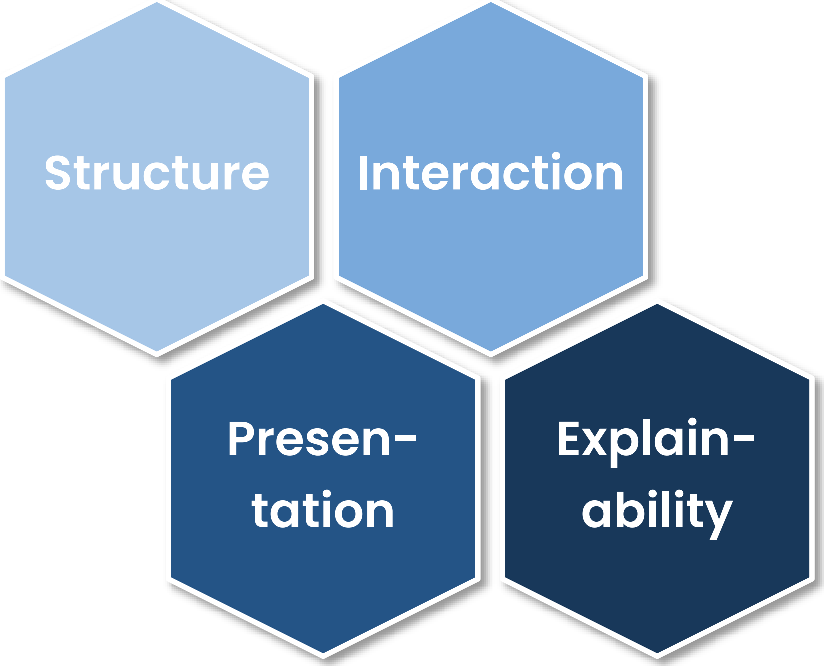
Biomedical AI Heuristics

Requirements

- Generalizing to new interfaces
- Adapted to the domain
- Easy to understand and apply

Approaches

- Adapting existing heuristics
- Reviewing literature of the designated domain
- Expert interviews



Biomedical AI Heuristics

Structure

1 Streamline main task
Lieberman (2009)

2 Provide full control
Amershi (2014), Gill (1996)

3 Orientation and Consistency
Nielsen (1990)

Interaction

4 Guide attention
Horvitz (1999), Sutton (2020)

5 Provide comparisons
Expert Interviews

6 Show impact
Jameson (2009)

7 User over system
Nielsen (1990), Horvitz (1999)

Always allow the user to correct errors by the AI efficiently and even turn off the AI completely if needed.

Presen-tation

8 Familiar language
Rzepka (2018)

9 Precise language
Nielsen (1990)

10 Familiar presentation styles
Expert Interviews

11 Appeal
Expert Interviews

Explain-ability

12 Explain data
De Fauw (2018)

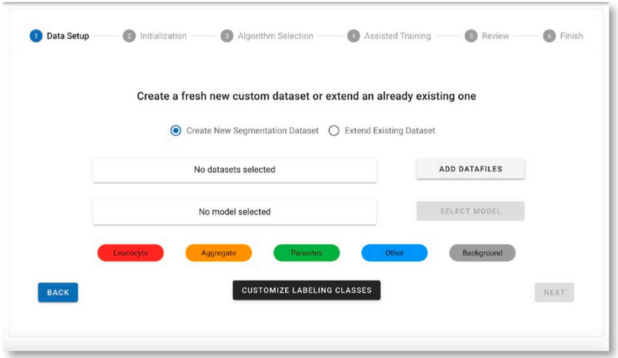
13 Explain process steps
Holzinger (2017)

14 Explain reasoning
Holzinger (2017)

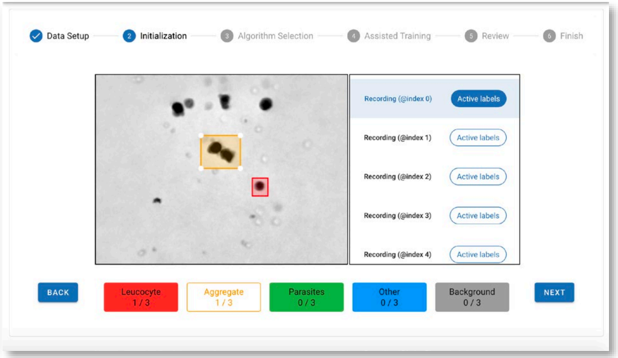
15 Highlight strengths and limitations *EU (2019)*

Interface Prototype

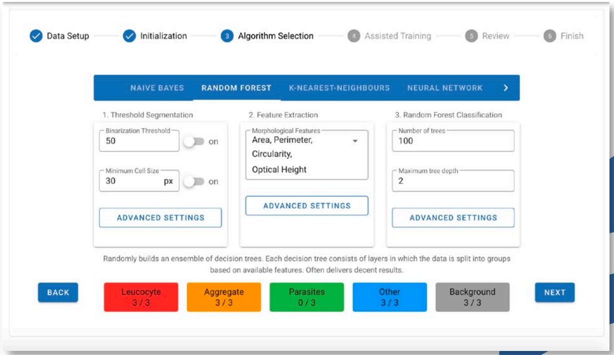
View 1: Data Setup



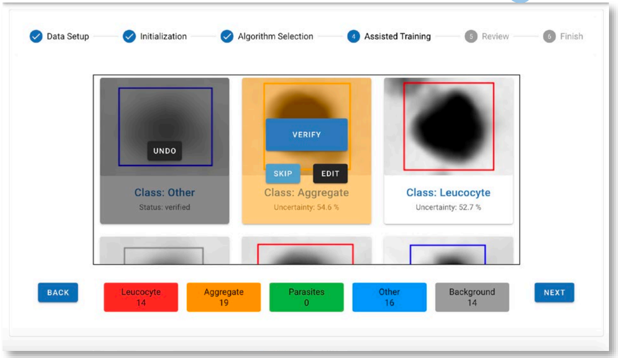
View 2: Initialization



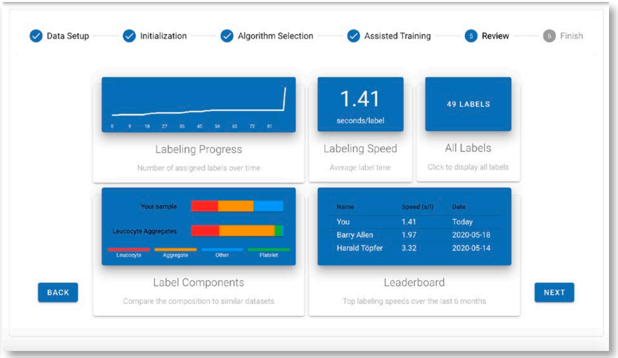
View 3: Algorithm Selection



View 4: Assisted Training

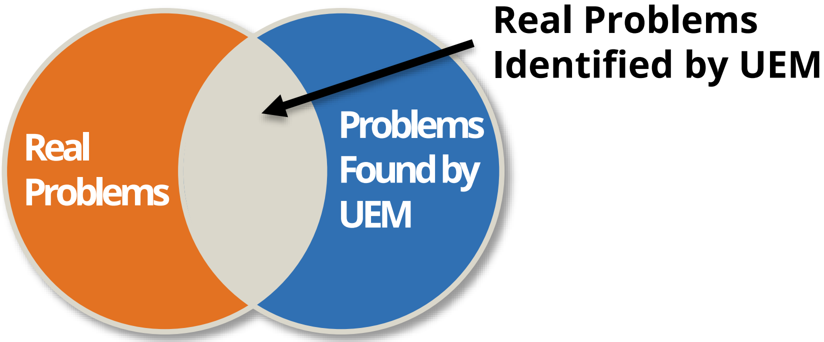


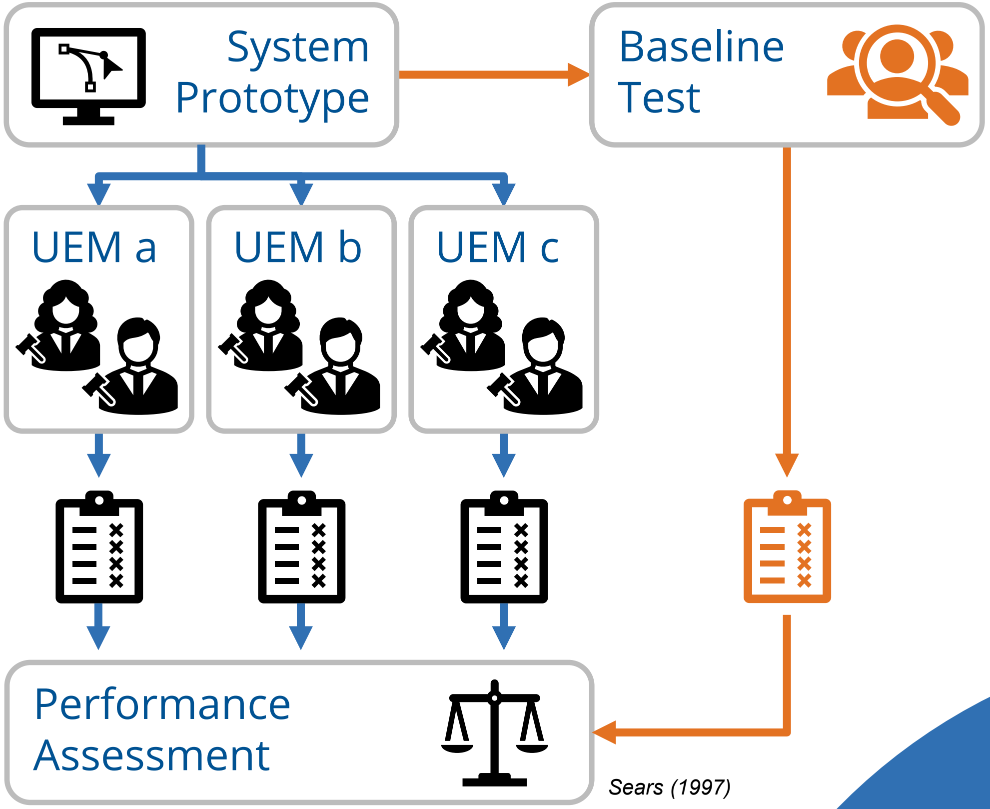
View 5: Review



Comparison of Usability Evaluation Methods

- Evaluate the interface with each Usability Evaluation Method (UEM)
- Conduct a baseline test to create a ground truth of usability problems
- Compare the performance using suitable metrics

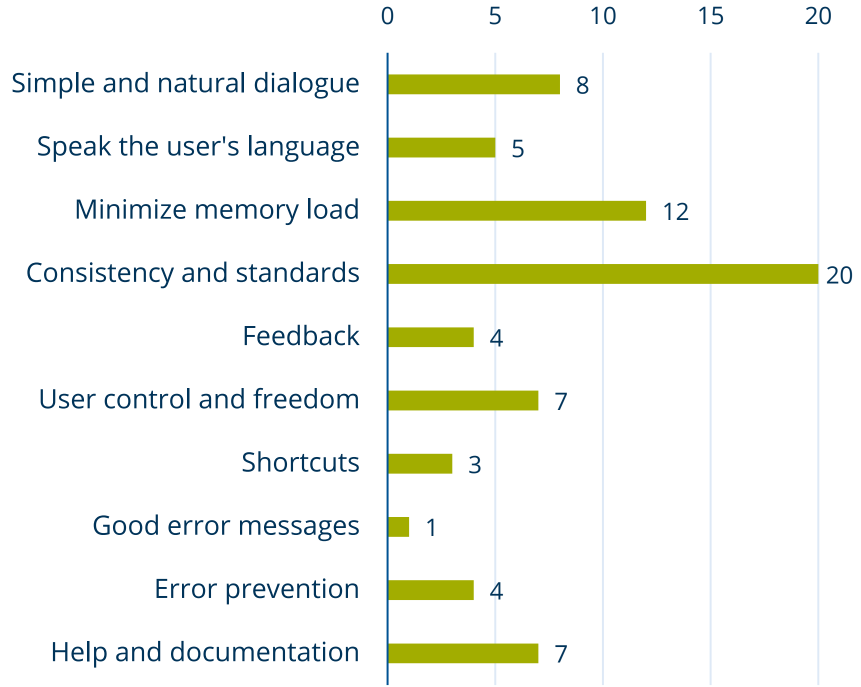




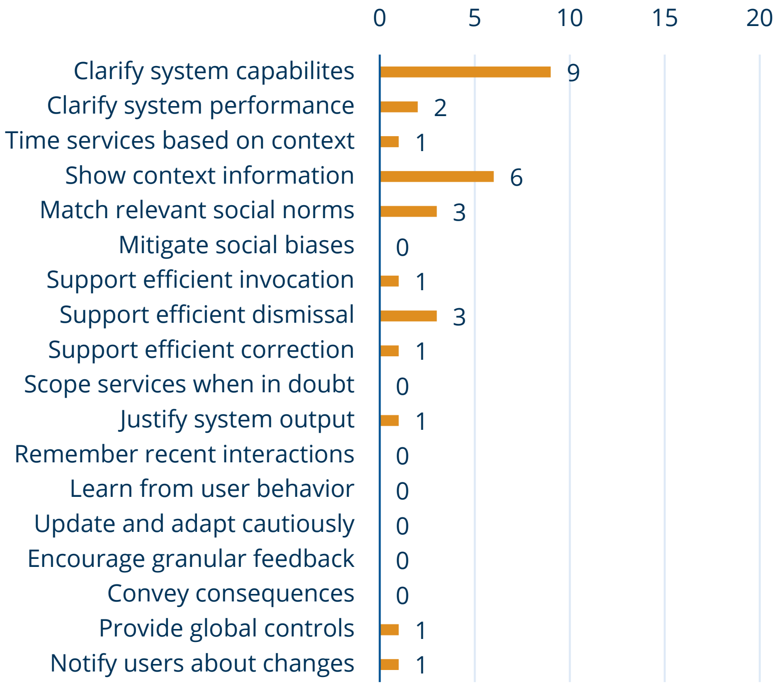
Sears (1997)

Heuristic Evaluation

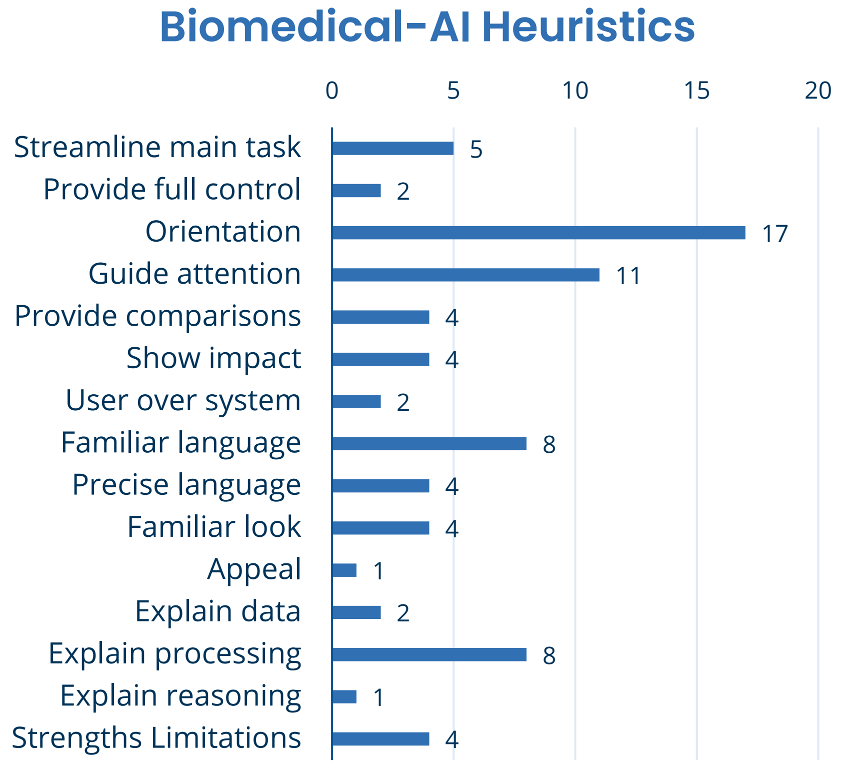
Nielsen's heuristics



Human-AI heuristics



Heuristic Evaluation

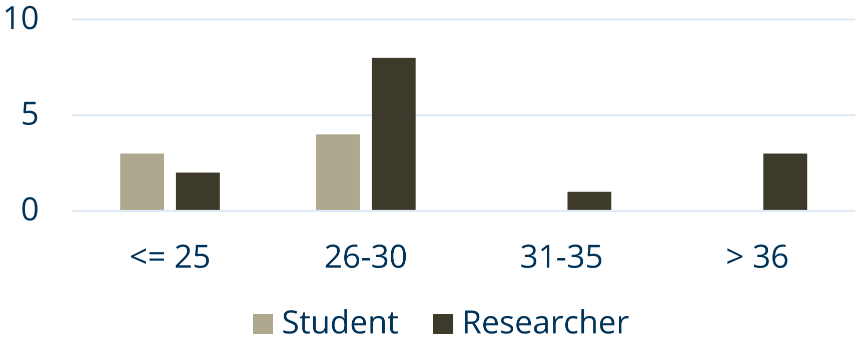


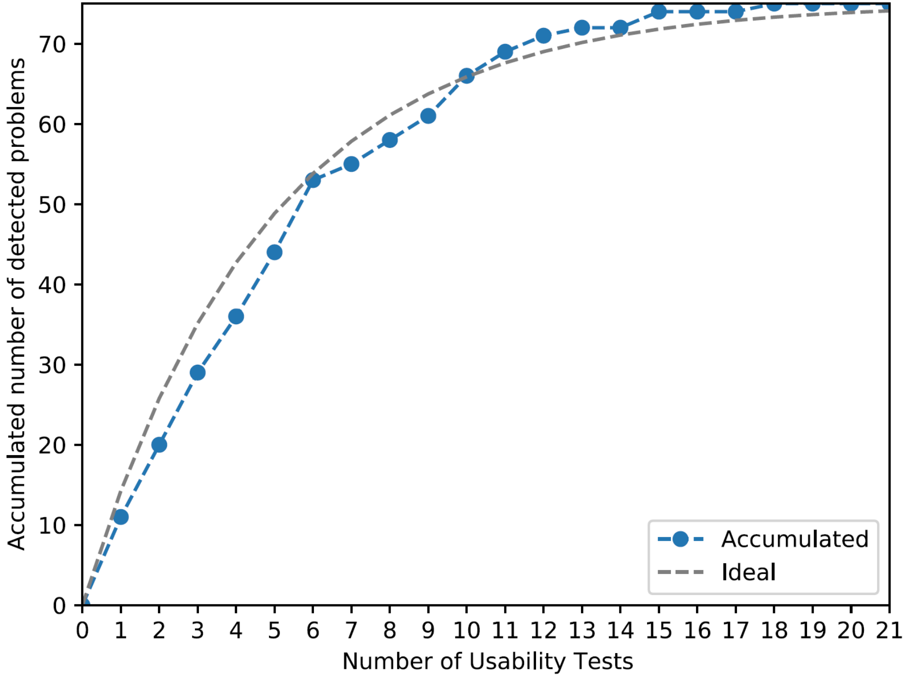
Heuristics	Predicted Problems
Nielsen's Heuristics	60
Human-AI interaction heuristics	26
Biomedical-AI heuristics	55

- Highest number of violations found by Nielsen's set of heuristics
- Experts stated that Human-AI heuristics were harder to apply

Empirical User Test

Demographics





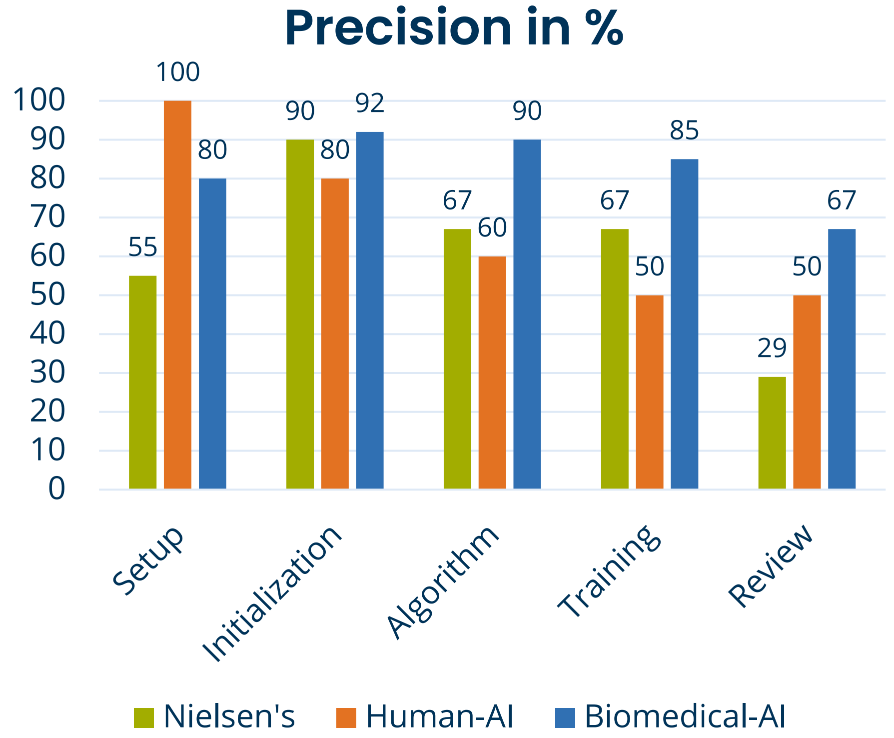
Lewis (1994), Hartson (2001)

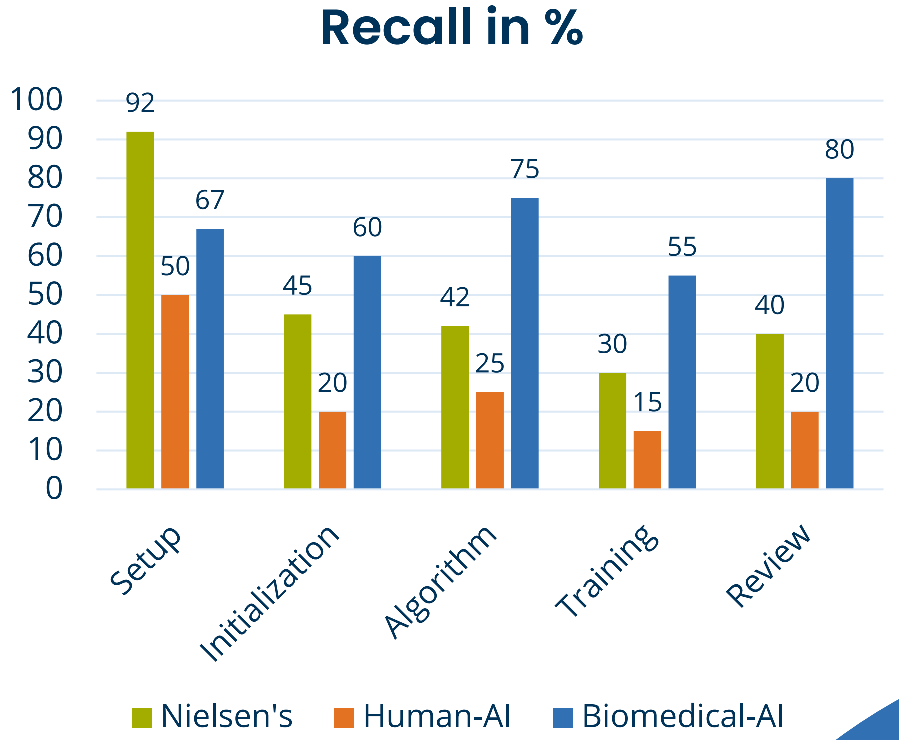
- Discovered 75 real usability problems
- Many have already been predicted by heuristics

Performance Comparison

Metric	Nielsen's heuristics	Human-AI heuristics	Biomedical-AI heuristics
Precision (Validity)	63.3%	73.1%	**85.5%**
Recall (Thoroughness)	50.1%	25.3%	**62.7%**
Average Severity	2.66	**2.84**	2.74
Weighted Recall	54.0%	28.9%	**69.0%**
Recall (Severity = 3)	38.9%	27.8%	**77.8%**
Recall (Severity = 4)	73.3%	46.7%	**93.3%**

Performance Comparison





Conclusion

Nielsen's heuristics

- Found most genuine problems
- Less efficient AI induced parts
- Reduced precision

Human-AI heuristics

- Worst precision and recall
- Some rules were not effective
- Difficult to apply by experts

Biomedical-AI heuristics

- Satisfying overall performance
- Uncovers most severe problems
- Best adaption to AI in this domain

Overall Remarks

- Demonstrated need for domain specific sets of heuristic rules
- Unweighted recall of 63% is helpful for further improvement
- Further evaluation on different interfaces needed

Thank You

References

S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, vol. 35, no. 4, pp. 105–120.

S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen et al. (2019). "Guidelines for human-AI interaction. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,* pp. 1–13.

J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin et al. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, vol. 24, no. 9, pp. 1342–1350.

European Commission (2019). High-Level Expert Group on Artificial Intelligence: Ethics Guidelines for Trustworthy AI.

T. G. Gill (1996). Expert systems usage: Task change and intrinsic motivation. *Management Information Systems Quarterly*, pp. 301–329.

H. R. Hartson, T. S. Andre, and R. C. Williges (2001). Criteria for evaluating usability evaluation methods," International Journal of Human-Computer Interaction, vol. 13, no. 4, pp. 373–410.

References

A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell (2017). What do we need to build explainable ai systems for the medical domain? *arXiv preprint arXiv:1712.09923*.

E. Horvitz (1999). Principles of mixed-initiative user interfaces. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 159–166.

A. D. Jameson (2009). Understanding and dealing with usability side effects of intelligent processing," *AI Magazine*, vol. 30, no. 4, pp. 23–23.

J.K. Kies, R.C. Williges, and M.B. Rosson (1998). Coordinating computer-supported cooperative work: A review of research issues and strategies. *In Journal of the American Society for Information Science*, vol. 49, no. 9, pp. 776–791.

J. R. Lewis (1994). Sample sizes for usability studies: Additional considerations. *Human factors*, vol. 36, no. 2, pp. 368–378.

H. Lieberman (2009). User interface goals, AI opportunities. *AI Magazine*, vol. 30, no. 4, pp. 16–22, 2009.

J. Nielsen and R. Molich (1990). Heuristic evaluation of user interfaces. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 249–256.

References

J. Nielsen and T. K. Landauer (1993). A mathematical model of the finding of usability problems. *In Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, pp. 206–213.

J. Nielsen (1995). Severity ratings for usability problems. *Papers and Essays*, vol. 54, pp. 1–2.

J. Nielsen (2005). Ten usability heuristics. (accessed 2022.12.17). [Online].

J. Noyes and C. Baber (1999). User-centred design of systems. *Springer Science & Business Media*.

C. Rzepka and B. Berger (2018). User interaction with AI-enabled systems: a systematic review of is research. *In Thirty Ninth International Conference on Information Systems*, vol. 39, pp. 1–17.

A. Sears (1997). Heuristic walkthroughs: Finding the problems without the noise. *International Journal of Human–Computer Interaction*, vol. 9, no. 3, pp. 213–234.

R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. *Nature Partner Journals Digital Medicine*, vol. 3, no. 1, pp. 1–10.

Biomedical Research AI Heuristics

Structure

1	Streamline main task	Make the system easy to learn. Focus on the main task that a system was created for and move unimportant side tasks away from the center of attention. The presentation of the main task should be simple to not overwhelm novice users.
2	Provide full control	Provide global controls and access to important model parameters that may be adjusted to boost the system performance. Complex tasks require fine tuning and advanced users often want to make these changes on their own.
3	Orientation and Consistency	AI systems are often complex which causes users to get lost. Always show the user where he is, what is currently going on and what he can do next. Navigation and interaction items should be easily noticeable and distinguishable.

Biomedical Research AI Heuristics

Interaction

4	Guide attention	It is easy for the user to lose track what is relevant in a complex AI system. Thus, it is important to guide the user's attention to where his action is needed. Keep him focused on his task and do not throw too many alarms and notifications at him.
5	Provide comparisons	Let biomedical users compare among similar data or among parameters when they need to judge an outcome of the system or make a decision on their own.
6	Show impact	The user needs to see how his actions and decisions directly influence the system and its performance.
7	User over system	Allow the user to correct errors of the AI efficiently at all times and even turn off the AI completely if needed.

Biomedical Research AI Heuristics

Presen-tation

8	Familiar language	Use non-technical language if possible. Pay attention to use the correct terminology for medical concepts.
9	Precise language	Avoid ambiguous wording for labels and commands that could trigger confusion in the user regarding how to do a task or how something works.
10	Familiar presentation styles	Do not use ways of presentation for the interface that biomedical users do not know from other interfaces they work with and that they cannot interpret.
11	Appeal	AI interfaces should have an appealing and premium look. This gives the user a feeling of using a state-of-the-art, high quality and thus trustworthy product.

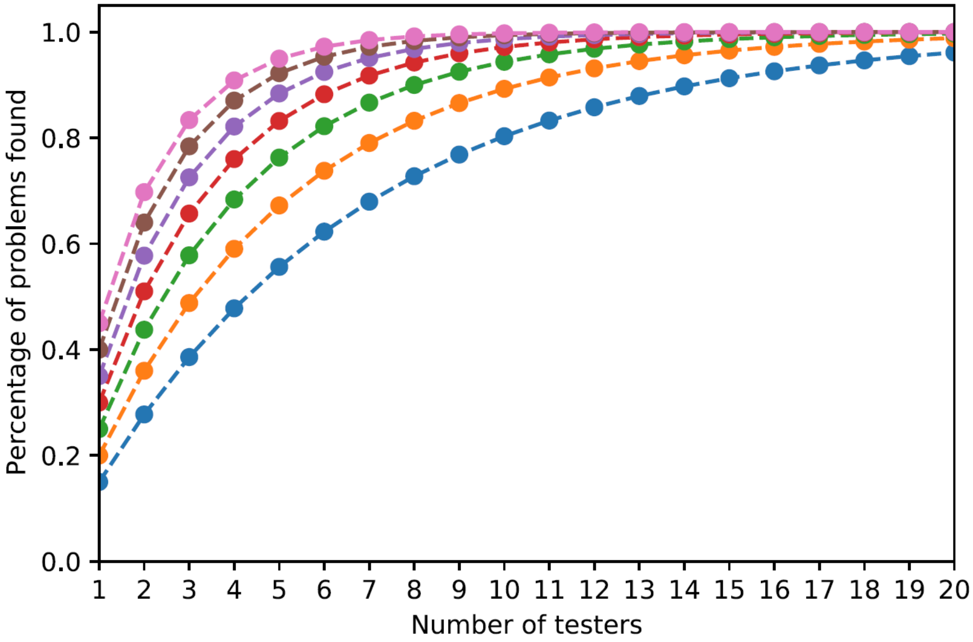
Biomedical Research AI Heuristics

Explain-ability

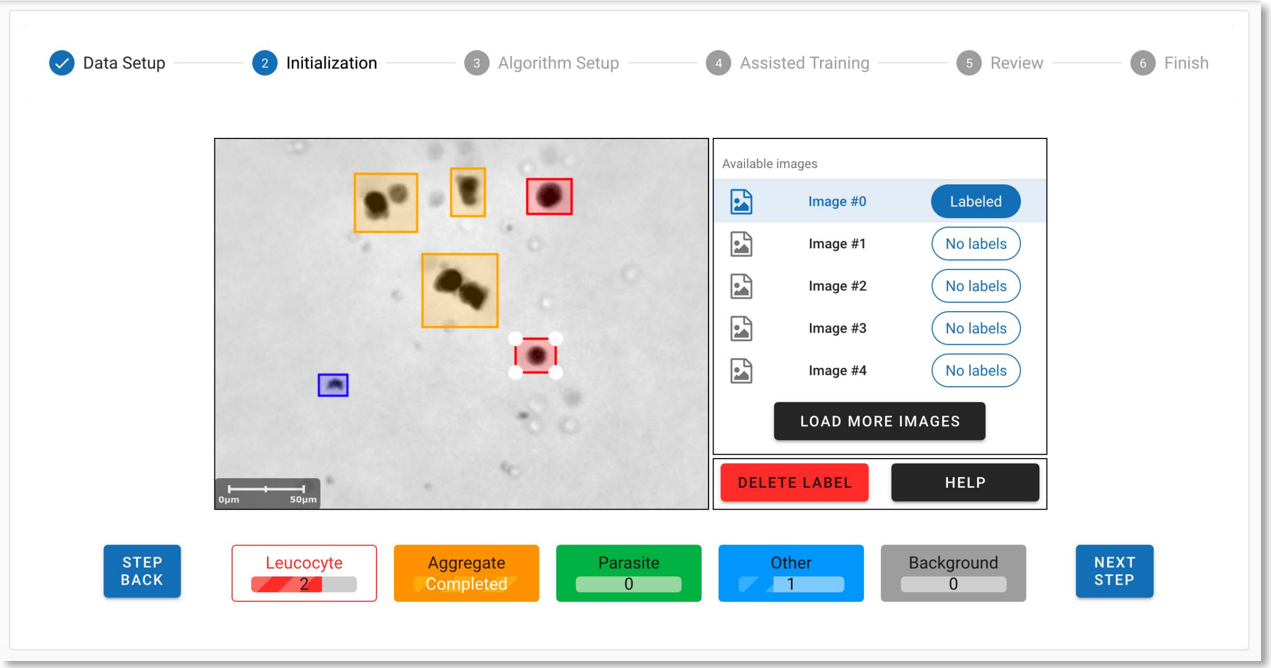
12	Explain data	The data that a task is based on needs sufficient explanation. The user has to develop an understanding of how it can be interpreted and how it is different from other data he may be familiar with.
13	Explain process steps	There needs to be a high-level explanation for the overall procedure that is performed by the system.
14	Explain reasoning	There has to be an explanation why and how the system derived a certain result or prediction.
15	Highlight strengths and limitations	Show what the strengths and weaknesses of the system are and what expectations are realistic. Highlight situations when the system may be underperforming or is in doubt.

Asymptotical User Testing

- Discovery of usability problems can be modeled as a Poisson process
- Number of detected usability problems reaches an asymptotic maximum
- Test with 21 users with a medical, biological and chemical background (7 students and 14 researchers)
- Worst Case Assumption



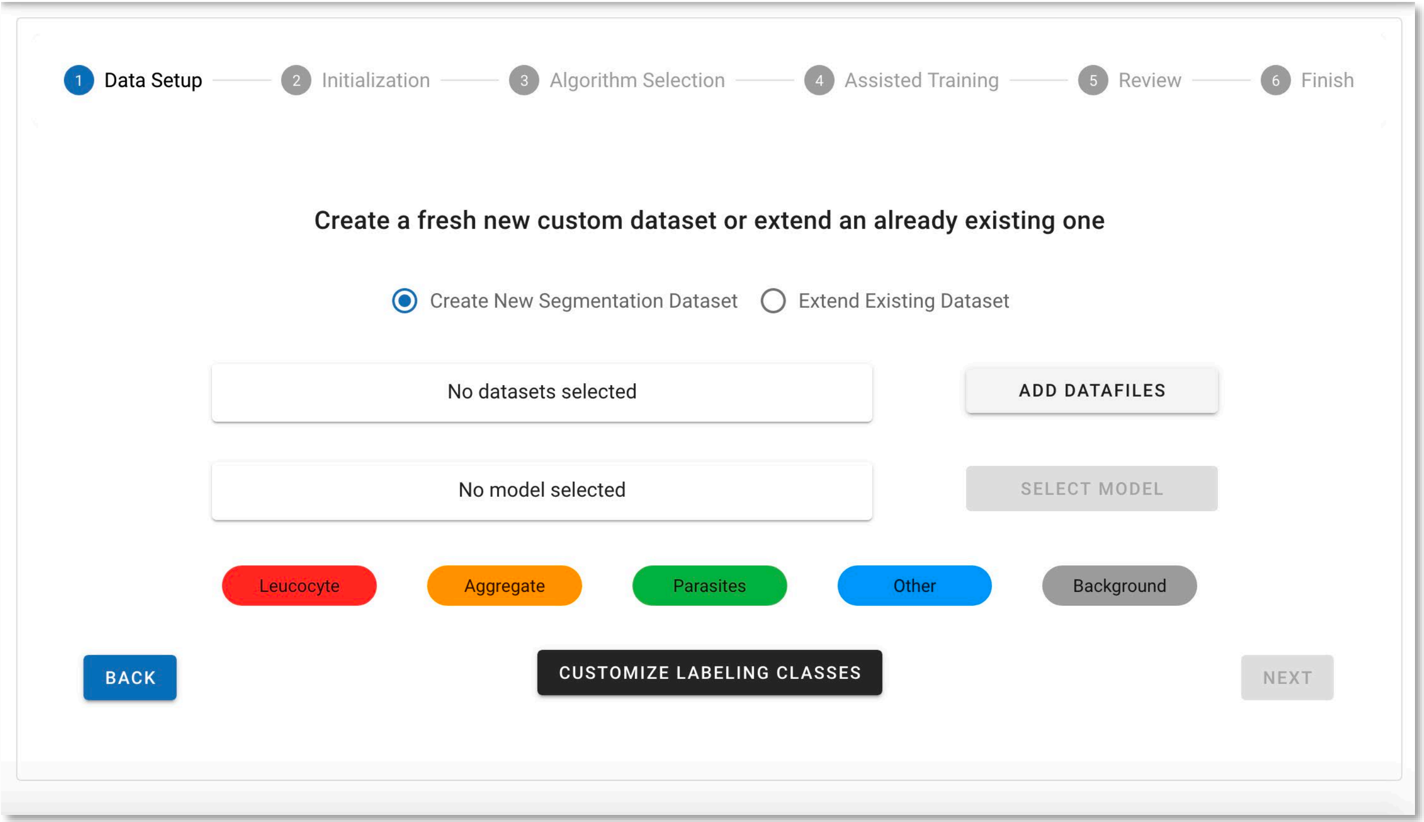
Interface Prototype



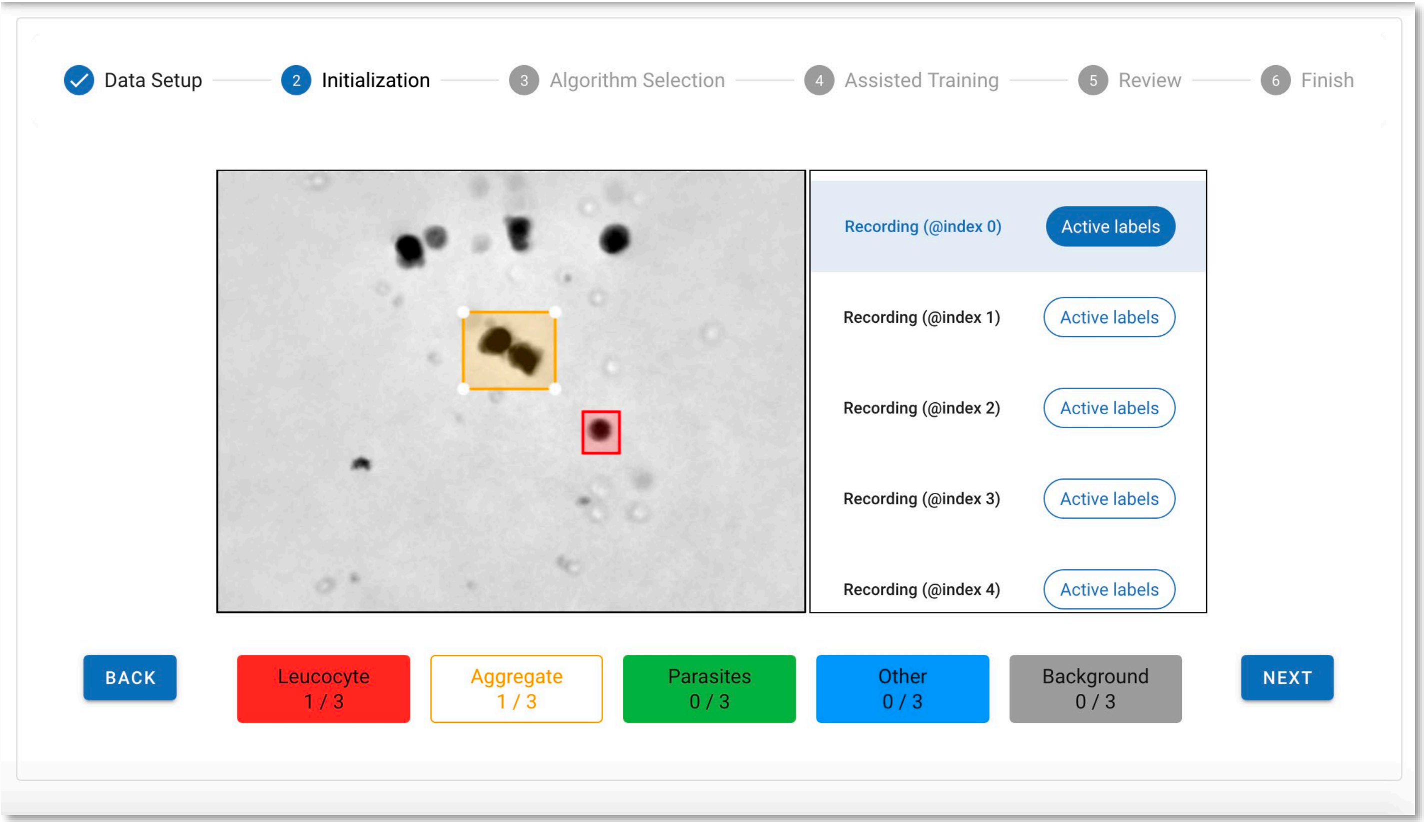
Human Assisted Labeling Tool

- Biomedical cell segmentation and classification
- Medical user and machine learning algorithm work hand in hand

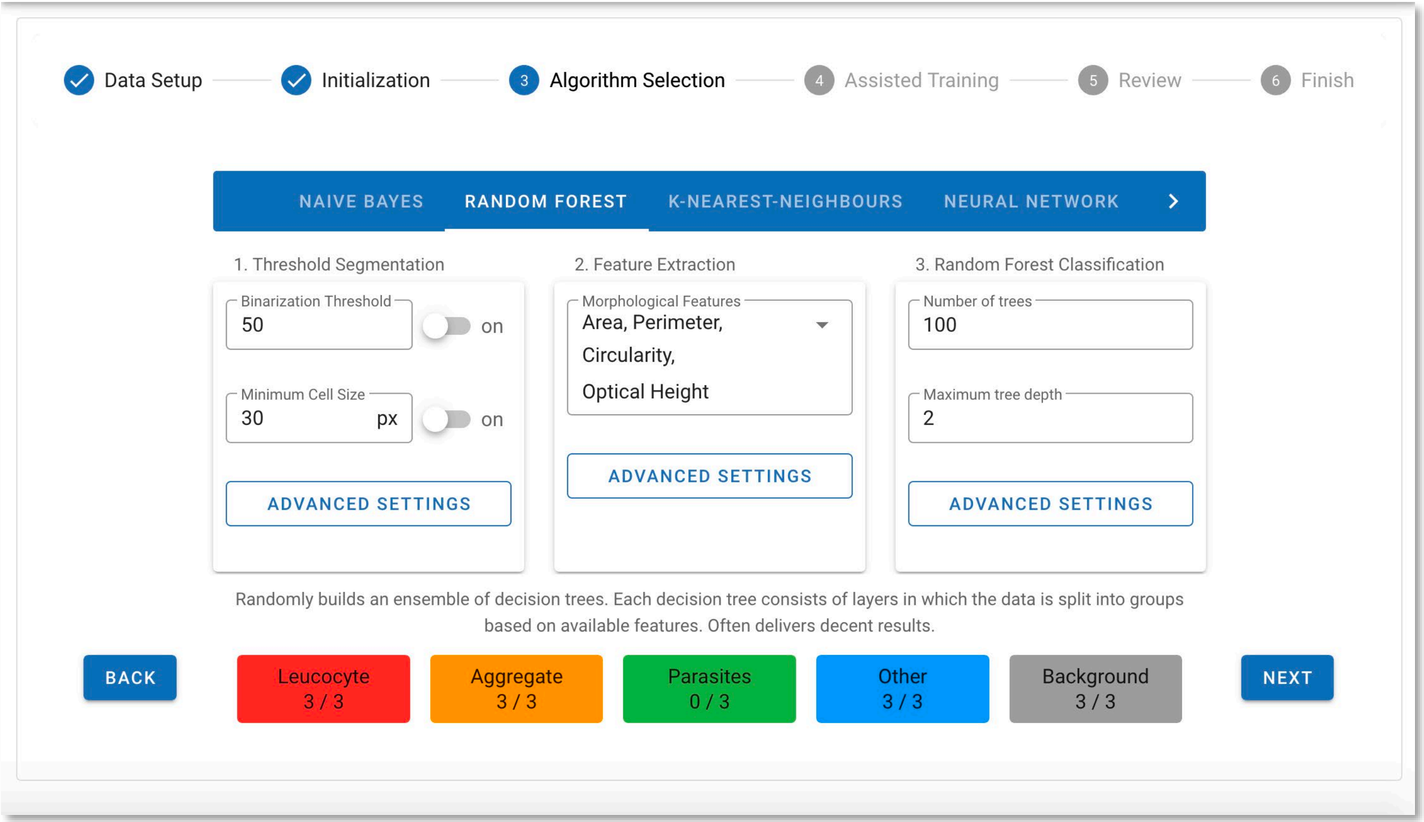
View 1: Data Setup



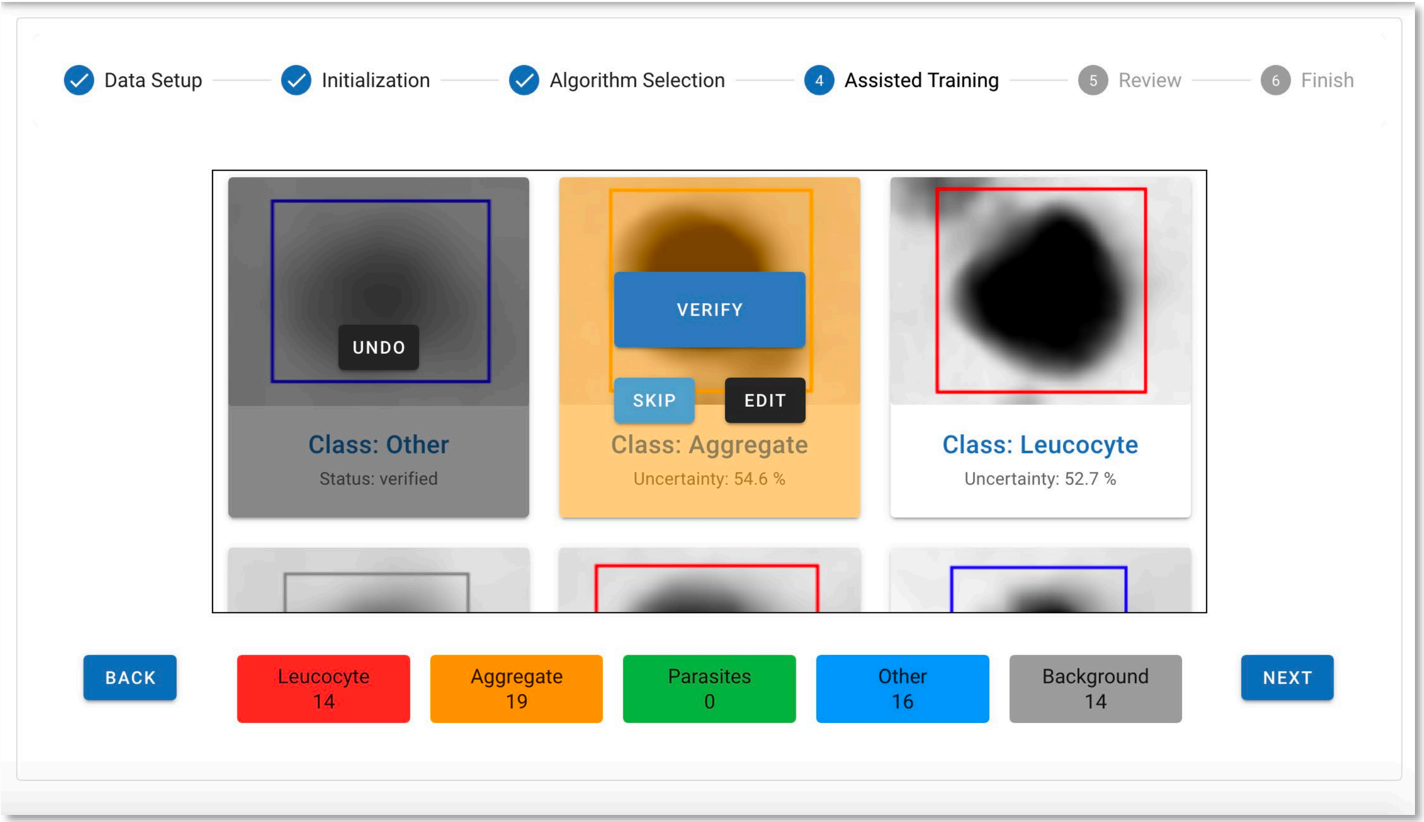
View 2: Initialization



View 3: Algorithm Selection



View 4: Assisted Training



View 5: Review