Optimizing AI software for automated PPD & CCI Protection

Honz Slipka, MSc
Senior Transparency Specialist

JULY 25TH, 2024

Agenda

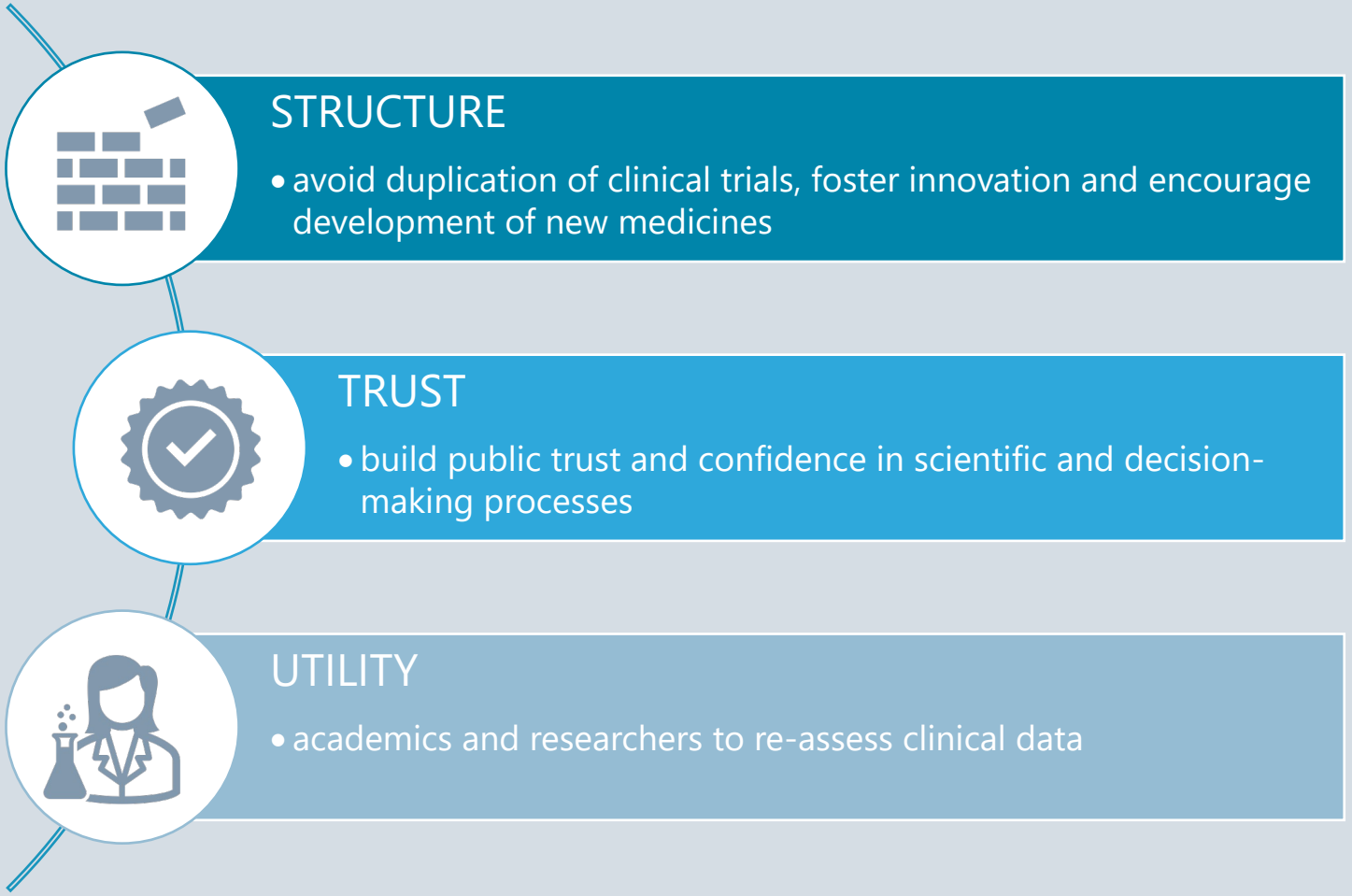
1. Objectives
2. Background
3. Examples
4. Implementation

Objective

PROACTIVE DISCLOSURE WITHOUT EXPOSURE

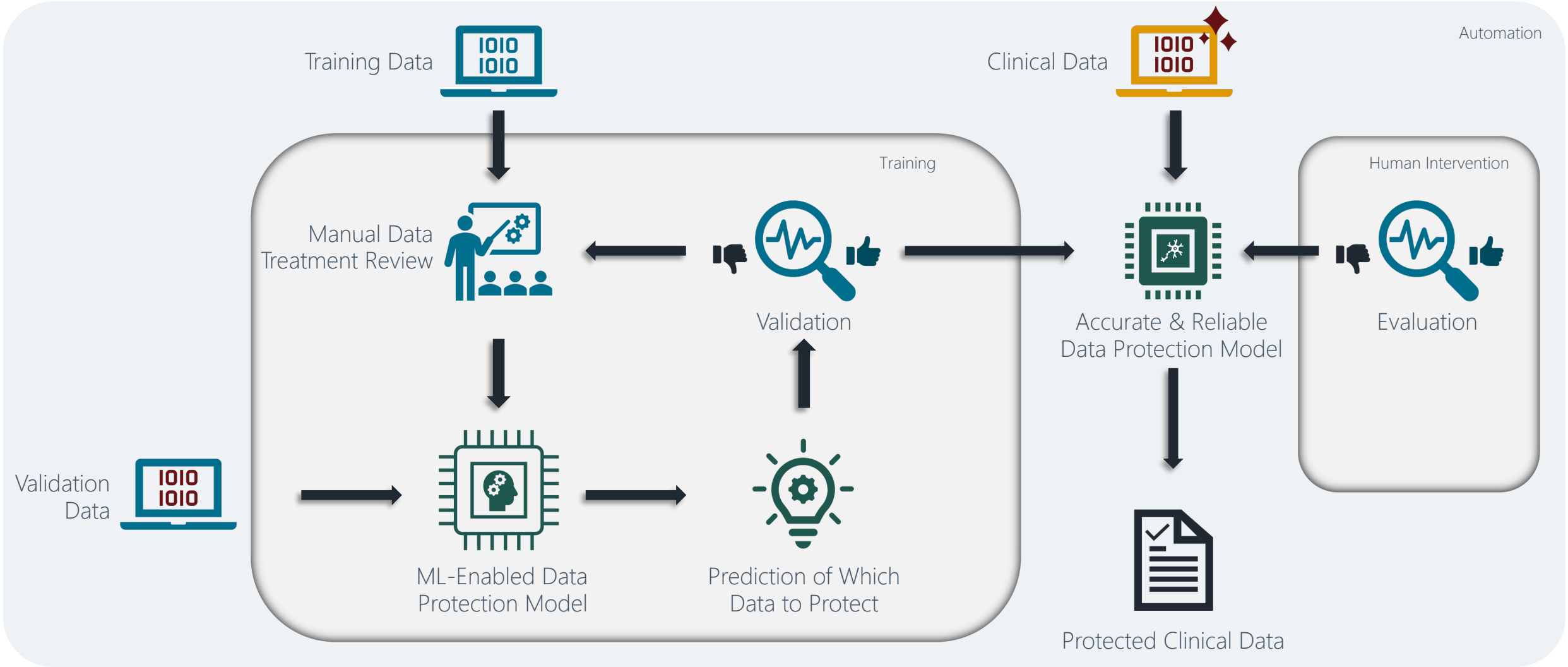
Clinical data to support pharmaceutical application for human medicines is submitted to regulatory publications for public disclosure.

- Personal data must be protected
- Submission must comply with regulations
- Emphasis should be on transparency



AI Data Protection

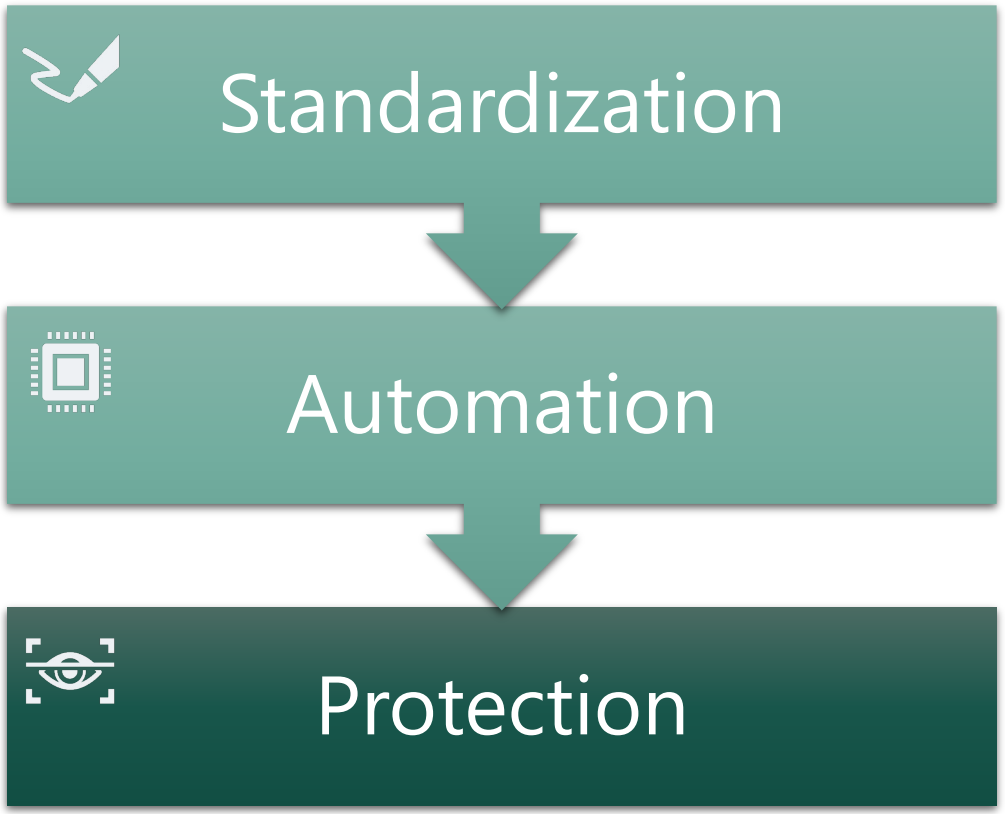
ML in the context of Clinical Data Protection

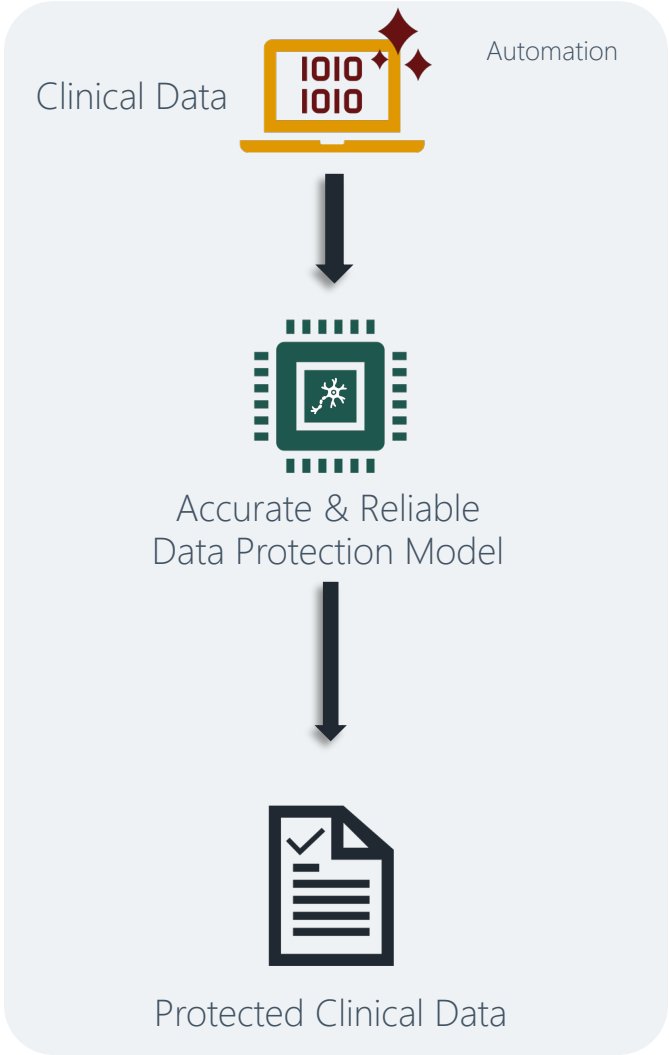


CERTARA

AI Data Protection

Focus on efficient protection of personal & commercial information





CERTARA

Data Protection

Focus on efficient protection of personal & commercial information

Standardization

- Harmonization of formatting for ease of understanding
- Template to expediate writing/linking

Automation

- Take advantage of AI, ML, and NLP tools
- Focus on future of technology and modernization of clinical research

Protection

- Provide the highest standard of patient privacy
- Focus on consistent treatment of direct/indirect patient identifiers

Dates

August 1st, 2023

Date	Day (01 to 1)	Month (08 to 8)	Month (Aug to August)	Year (2023 to 23)
•ddMMMyyyy (01AUG2023) •dd-mmm-yyyy (01-aug-2023) •dd-mm-yyyy (01-08-2023) •dd/MM/yyyy (01/08/2023) •MM-dd-yyyy (08-01-2023) •MM/dd/yyyy (08/01/2023) •1st of Aug •First of the month (ie dosing or sample dates) •"New years eve"	•dMMMyyyy (1AUG2023) •d-mmm-yyyy (1-aug-2023) •d-mm-yyyy (1-08-2023) •d/MM/yyyy (1/08/2023) •MM-d-yyyy (08-1-2023) •MM/d/yyyy (08/1/2023)	•d-M-yyyy (1-8-2023) •d/M/yyyy (1/8/2023) •M-d-yyyy (8-1-2023) •M/d/yyyy (8/1/2023) •dd-MM-yyyy (01-8-2023) •dd/M/yyyy (01/8/2023) •M-dd-yyyy (8-01-2023) •M/dd/yyyy (8/01/2023)	•ddMMMyyyy (01AUGUST2023) •dMMMyyyy (1AUGUST2023) •1st of August •August 1st	•20+ more variations

False Negative	False Positive	Error
•Missed data •New format, untrained software, etc.	•Incorrectly identified data •Overtraining & noise •Study ID **Jan-24-12** (a *hypothetical* Janssen study from 2024) could be identified as a date.	•Mismanaged data •08-01-23 •***dd-mm-yy*** or ***mm-dd-yy*** or ***yy-mm-dd***

CERTARA

Dates

August 1st, 2023

Standardization

- Harmonization of formatting to limit variation
- Template to expediate writing/linking

Automation

- AI relies on high quality input
- Focus on future of technology and optimization of current processes

Protection

- Focus on maintaining high standard of patient privacy
- Dates can be knowable

OMIT

unnecessary details

- Do not include times unless necessary
- Move towards using study day

Does not reveal sensitive event dates

Allows for time tracking of study events

Maintains data utility

LIMIT

appearance within the text

- Do not include in sections other than narratives
- Only include where study day cannot be used (birth date etc.)

COMMIT

to harmonization of text

- Use standardized format: **yyyy-mm-dd** (ISO)
- Consistent across all studies & documents

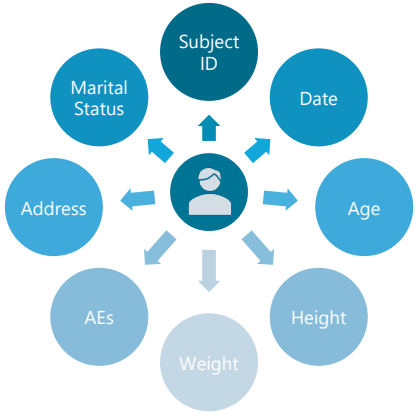
CERTARA

Implementation

Omit

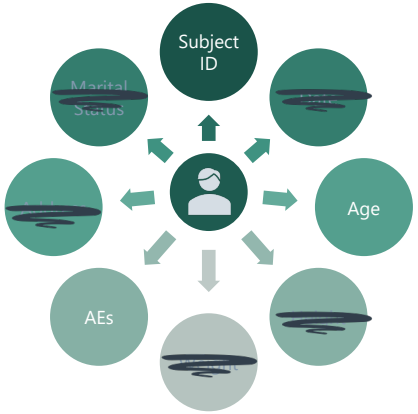
Current State

- Unnecessary amount of information
- Model need to deal with noise
- More iterations requiring more work and resources



Future State

- Standardized information quickly searchable
- Models fine-tuned to specific data
- Less sensitive data, less work

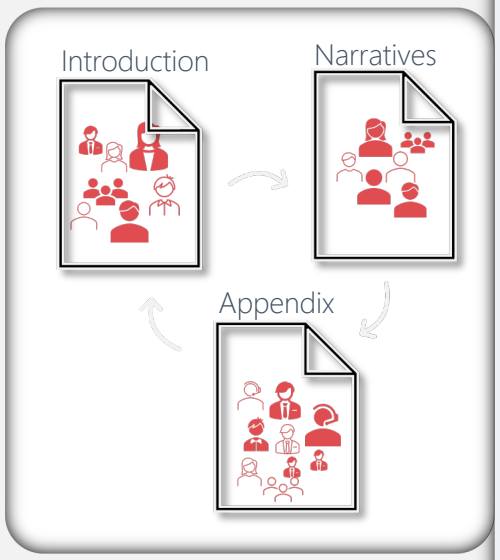


Implementation

Limit

Current State

- Data distributed throughout entire study
- Model must be applied to entire study
- More review requiring time and resources



Future State

- Information centralized to specific location
- Model can be targeted to particular sections
- Less input data to feed into models

Implementation

Commit

Current State

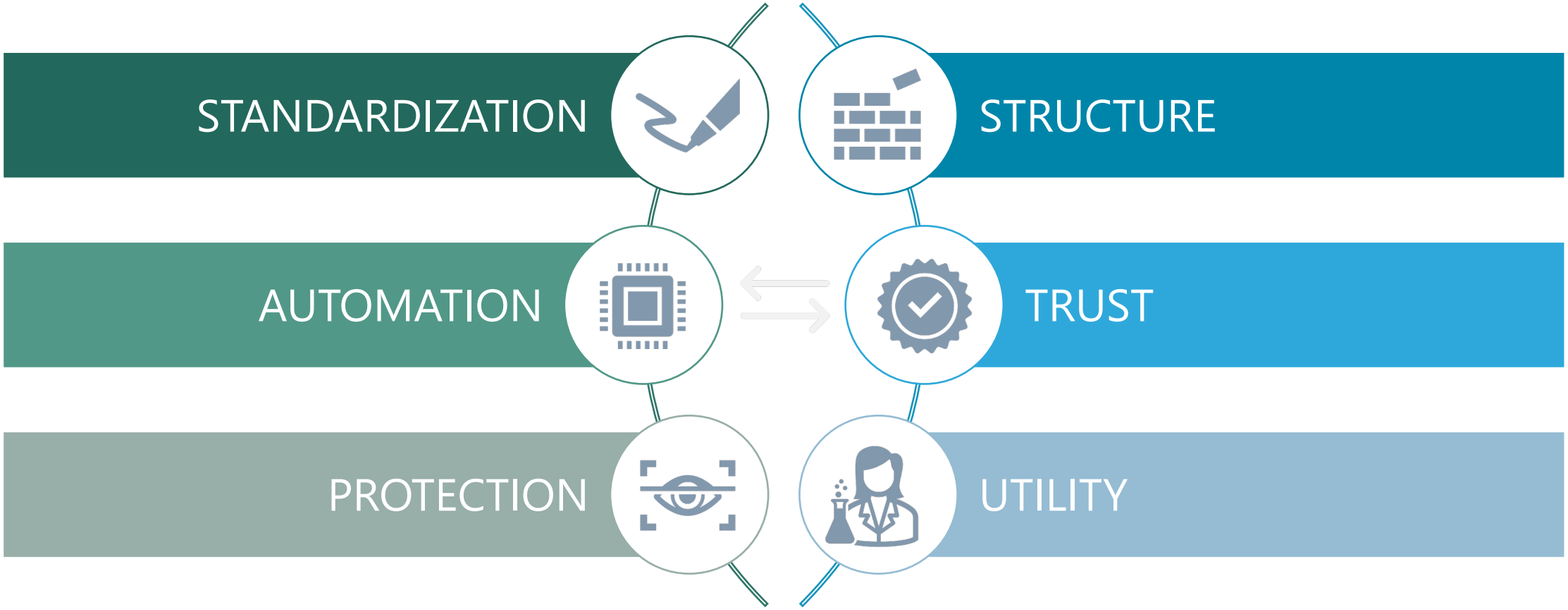
- Excessive variability in data
- Model must be highly sensitive
- Typically under-redaction with false negatives

Future State

- Variations of data simplified
- Model can be aimed at specific formatting
- Less likelihood of false negatives

Data Protection

Good clinical authoring helps automate and data protection



Honz Slipka, MSc
Senior Transparency Specialist

THANK YOU