

Trust in a Safeguards Voice User Interface for a Nuclear Material Measurement Task

Zoe N. Gastelum*, Kristin M. Divis, Breannan C. Howell, Jamie L. Coram
Sandia National Laboratories

July 2024
INMM Annual Meeting | Portland, OR
*zgastel@sandia.gov

Safeguards inspections take place in **difficult environments**:

- Loud environments
- Jetlag
- Language barriers
- Time pressure
- Potential safety hazards



Image: IAEA



Digital Assistants for Nuclear Safeguards Inspectors

- Completing inspection paperwork
- Accounting for the presence of nuclear material containers with a barcode reader
- Inspector wayfinding in nuclear facilities
- Seal examination
- Physical design and layout measurements
- Robotic inspection assistants
- Identification and quantification of safeguards-relevant isotopes
- Review of safeguards surveillance footage

Appropriate trust in the system:

too much – complacency, lose benefit of expertise
too little – lose performance benefits, smaller ROI



How to provide information in order to optimize trust?

Model performance and industry standards

- + Safeguards:
 - Confidence (x2)
 - Explainability
 - Provenance
 - Granularity

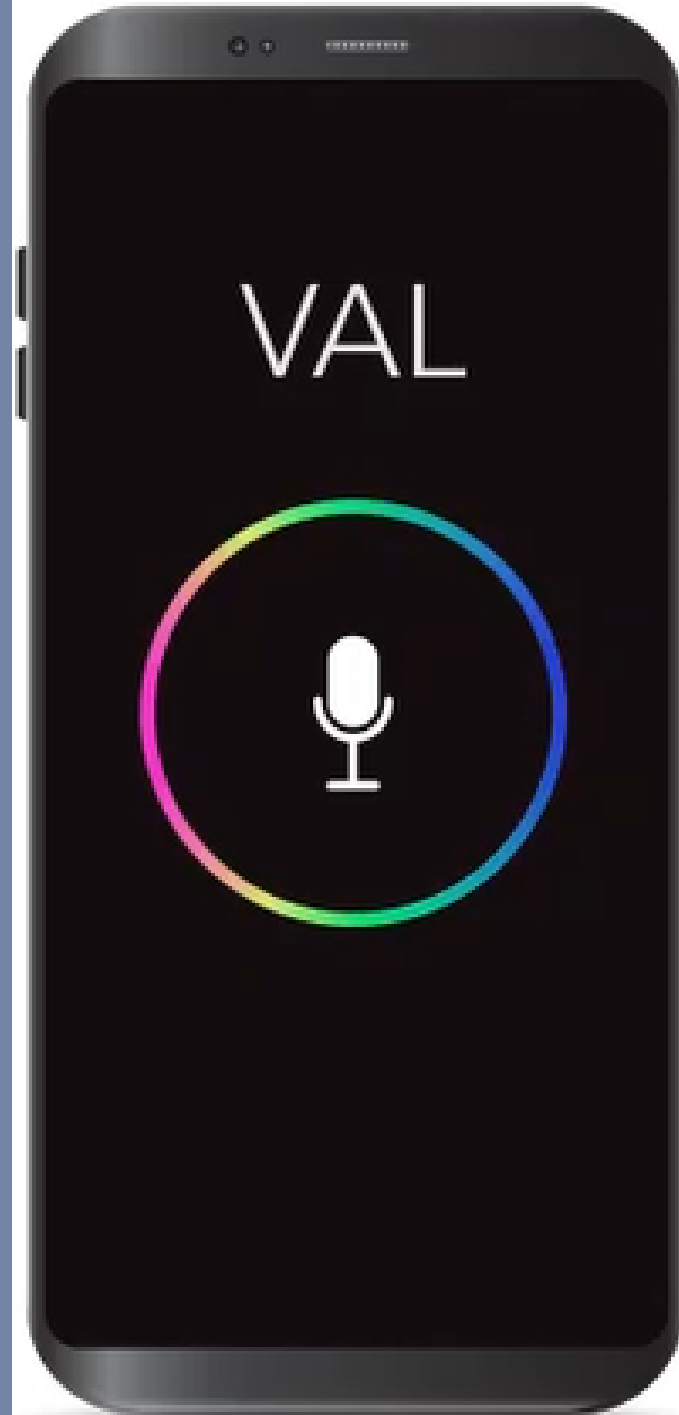




We measure trust via **human performance** testing with simulated Voice User Interface.

- Post-task Questionnaire
- Quantitative behavioral data
- Trial-by-trial feedback on user confidence

Meet VAL, our simulated digital voice assistant for safeguards inspectors.




1) Seal Examination Task

- VAL uses OCR to read seal ID
- VAL reports on predicted tamper status
- Participants view the seal and decide to keep seal, or replace (they can see the seal)
- Trust measures: Confidence (2 ways) and explainability



Seal confidence (numerical)




“Seal 46184 shows signs of tamper, with 70% confidence”

Do people respond more trustfully to a system that has “medium” confidence than a system that has “64.8% confidence”?

- No confidence information
- Low-detail (“low”, “medium”, “high”)
- Moderate-detail numerical (“65% confidence” or “70% confidence”)
- High-detail numerical (“64.8% confidence” or “64.9% confidence”).

Seal confidence (rerun)



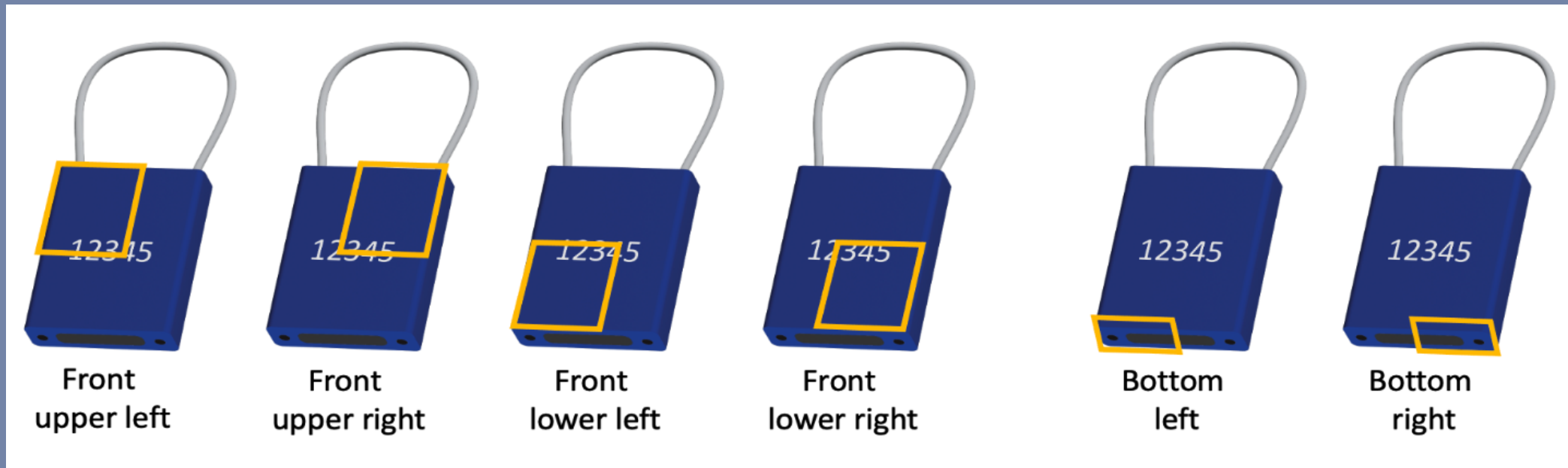
“Seal 46184
shows signs of
tamper, Re-run
suggested.”

Manipulated the system's response to confidence information.
VAL **suggested** or **required** a re-run for low confidence trials. Other trials were assumed to have an acceptable level of confidence

Seal – explainability

VAL provided information about the location of the tamper detection

- No location information
- Moderately-detailed (front, bottom).
- Highly-detailed (quadrant – “front lower left”)



Seal Examination Results

Participants who received **explainability** information or re-run suggestions **had higher levels of trust** and reliability in VAL

For **confidence** (numerical) experiments, there was **no significant difference** in trust and reliability reporting

- When VAL said there were signs of **tamper but with low confidence levels**, had **smaller increases in response time** than the control group -> helping participants dismiss the incorrect information faster
- Higher accuracy of participants receiving low confidence indicators from VAL when she provided an incorrect answer.



2) Material measurement

- VAL uses OCR to read container ID
- VAL reads information from measurement equipment
- Participants compare to declaration list
- **Trust Measures:** Confidence (re-run), Provenance, and granularity



Confidence re-run

- Control (re-measure never available)
- Available (always available),
- Suggested (if the measurement was near the edge or far outside the acceptable range).

*Re-measure was not limited to a subset of trials and we did not test a required re-measure manipulation.

Container ID	Material	Declared Enrichment	Element weight (kg)	Number of Batches
60722	Natural uranium	0.8%	2137	6
51668	Enriched uranium E < 20%	5.2%	2331	10
92368	Enriched uranium E < 20%	4.4%	2448	5
10364	Natural uranium	0.8%	2035	5
60037	Depleted uranium	0.6%	2435	5
93185	Enriched uranium E < 20%	1.8%	2054	7
95809	Enriched uranium E < 20%	1.3%	2473	7
94071	Enriched uranium E < 20%	2.9%	2387	7
18150	Enriched uranium E < 20%	2.9%	2363	1
26835	Enriched uranium E < 20%	7.9%	2260	10



Provenance

Does historical information impact user trust?



“Container 46184 is easy to measure”

VAL included historical information about each container

- Prior measurement of the same container
- History of being easy or difficult to measure accurately

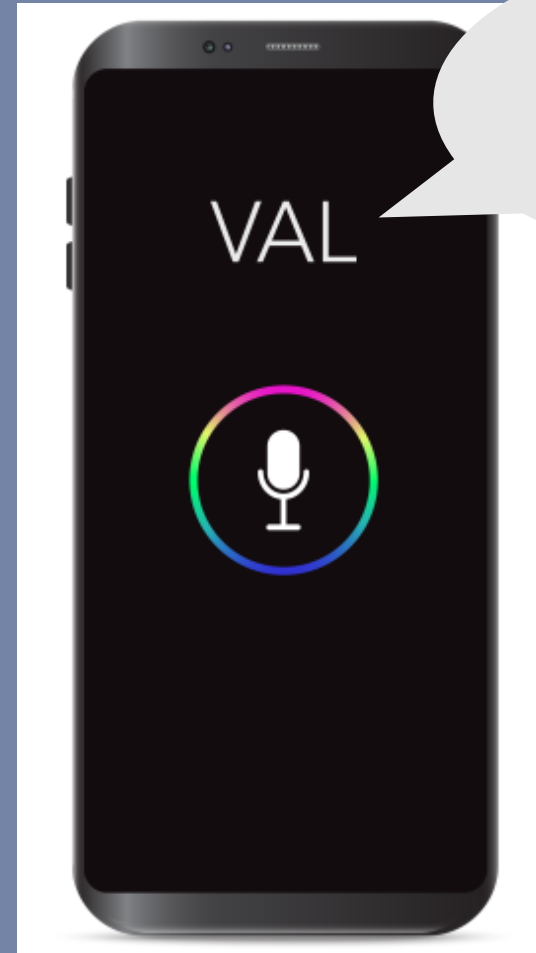
Participants completed trials for all three conditions.

Granularity

Does the number of significant digits provided impact user trust?

Using mass measurements enabled a shift in the placement of the decimal point without changing actual value:

- Participants saw one unit
- Different numbers of sig. dig.
- Confidence measured per trial



“Container
46184
measures
3.094 grams”

- 3 grams
- 3.1 grams
- 3.09 grams
- 3.094 grams

Material Measurement Results

Granularity – pending

Confidence/Re-run – no significant difference in trust or reliability.

- Behavioral data indicated re-run compliance
- Differences in response time and accuracy being explored.

Provenance - no large differences in trust/reliability.
Remainder of trust indicators are still being explored.



Observations

- Users prefer more information
- Correct information can support user assessment in trust/compliance (confidence).
- Other information can build trust, even if not impacting user performance (explainability)
- Users like options more than requirements (re-run), even if they behave the same for each



Image: IAEA, 2021

Future Work

- Publish findings and recommendations
- Explore visual + voice
- Other modalities – haptics, robotics...



Photo: Boston Dynamics, USA



Thank you!

This work was funded by the National
Nuclear Security Administration's Defense
Nuclear Nonproliferation Research &
Development, Safeguards Program Office.

