

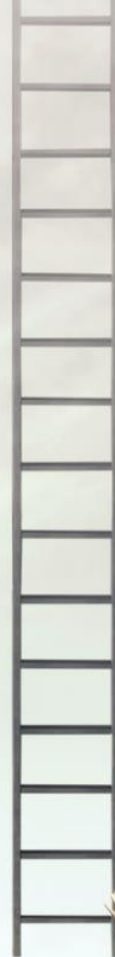
**Using Artificial Intelligence to identify  
inherent errors in medical data that  
even experts can't detect**

Dr Don Perugini  
Co-Founder

e: [Don@Presagen.com](mailto:Don@Presagen.com)  
l: <https://linkedin.com/in/donperugini>  
t: @DrDonPerugini  
w: [Presagen.com](http://Presagen.com)



SAN FRANCISCO | ADELAIDE | LONDON



# Medical data is inherently poor quality

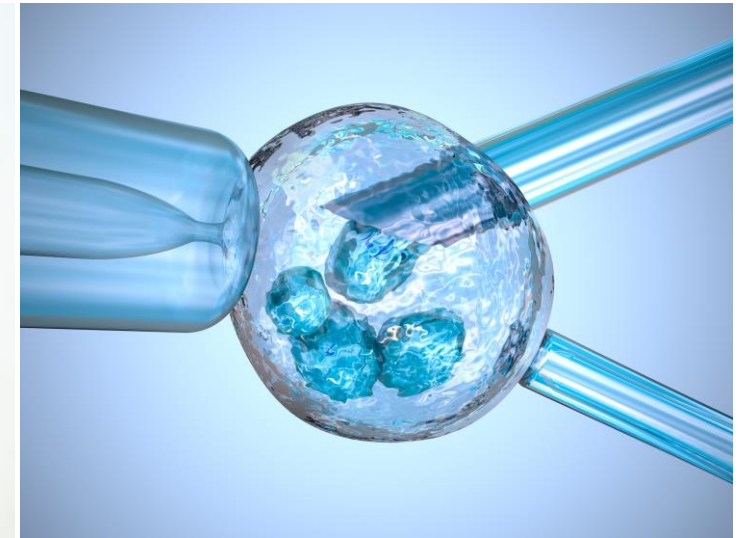
**Subjectivity**



**Uncertainty**



**Unknown  
Confounding Factors**



# Poor quality data negatively impacts many areas of healthcare and reliable treatment for patients

## Artificial Intelligence



## Precision Health



## Clinical trials



# Example: What if we wanted to identify symptoms associated with COVID patients?

COVID Patients



Non-COVID Patients



Symptom  
Data



Common cold and flu

# ... and what if Rapid Antigen Tests (RAT) were used to diagnose whether a patient has COVID?

COVID Patients



Non-COVID Patients



Symptom  
Data



RAT used to  
diagnose COVID



Common cold and flu

# The high error rate in the test will produce a large proportion of inherent errors in the data

COVID Patients



**>20% error in  
detecting COVID**

Non-COVID Patients



**<3% error in detecting  
Non-COVID**

Symptom  
Data



RAT used to  
diagnose COVID



# Experts cannot manually detect these inherent errors



**Many assume errors can be detected manually**

Other barriers to manual verification of data:

- *Scalability*: sheer scale of data to be reviewed
- *Privacy*: privacy regulation preventing manual review
- *Context*: lack of contextual information about the data





# Consider AI trained to detect cats and dogs

Cats



Dogs



# AI learns patterns and features from historical data to distinguish between cats and dogs

Cats



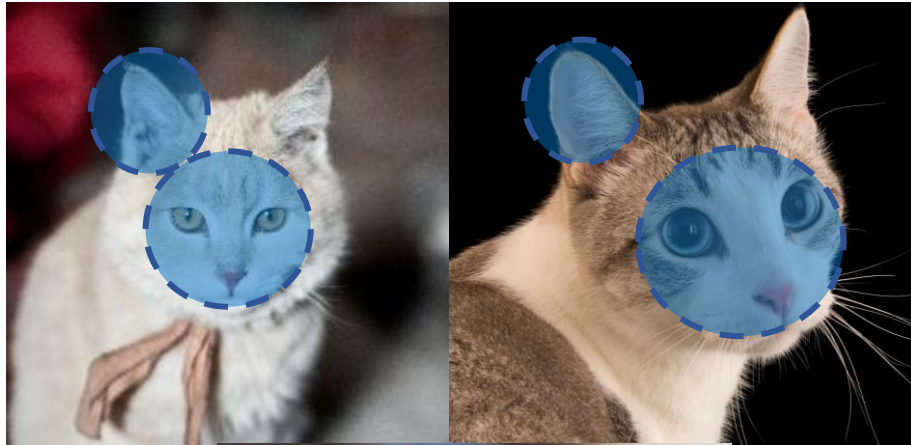
Dogs



Data needs to be **labelled** as 'Cat' or 'Dog' to train the AI

# Errors in data creates inconsistencies during AI training

Cats



Dogs



AI can't match the features that look different

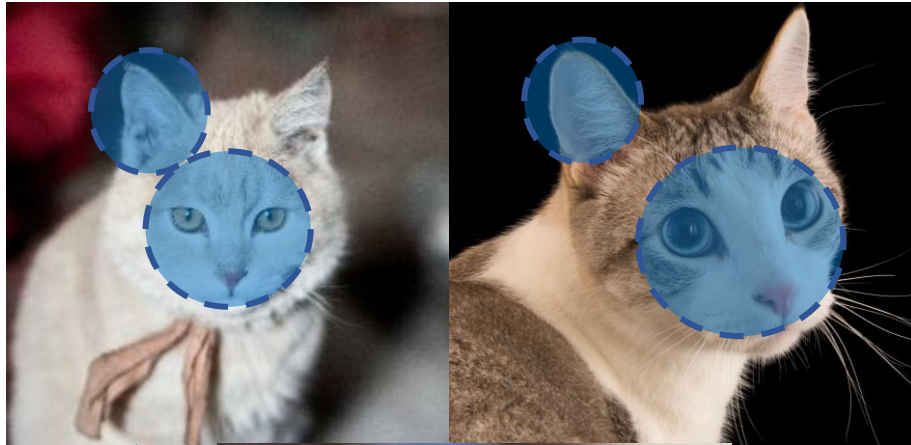
More often during training the AI will match the features associated with "Dogs" and think it is incorrect

**Cat??**

**Like trying to fit a square peg into a round hole!!**

# 'Untrainable' consistently incorrect data are likely errors

Cats



Dogs



~~Cat??~~

AI can't match the features that look different

More often during training the AI will match the features associated with "Dogs" and think it is incorrect

**Label 'Cat' is an error.  
Likely to be a 'Dog'.**

# How well does it work?

## Cats and Dogs



## Vehicles



## X-Ray images to detect pneumonia



## Embryo images to predict viability/pregnancy



## Non-Image Record Medical Data

id	patient	age	sex	date	time	type	value	unit	category	value	unit	category	value	unit	category
1	101	65	M	2010-01-01	10:00	BP	120	mmHg	BP	120	mmHg	BP	120	mmHg	BP
2	101	65	M	2010-01-01	10:05	HR	75	bpm	HR	75	bpm	HR	75	bpm	HR
3	101	65	M	2010-01-01	10:10	RR	18	breaths/min	RR	18	breaths/min	RR	18	breaths/min	RR
4	101	65	M	2010-01-01	10:15	SpO2	98	%	SpO2	98	%	SpO2	98	%	SpO2
5	101	65	M	2010-01-01	10:20	Temp	37.5	C	Temp	37.5	C	Temp	37.5	C	Temp
6	101	65	M	2010-01-01	10:25	Weight	70	kg	Weight	70	kg	Weight	70	kg	Weight
7	101	65	M	2010-01-01	10:30	Height	175	cm	Height	175	cm	Height	175	cm	Height
8	101	65	M	2010-01-01	10:35	Cholesterol	200	mg/dL	Cholesterol	200	mg/dL	Cholesterol	200	mg/dL	Cholesterol
9	101	65	M	2010-01-01	10:40	Blood Sugar	100	mg/dL	Blood Sugar	100	mg/dL	Blood Sugar	100	mg/dL	Blood Sugar
10	101	65	M	2010-01-01	10:45	Hemoglobin	15	g/dL	Hemoglobin	15	g/dL	Hemoglobin	15	g/dL	Hemoglobin
11	101	65	M	2010-01-01	10:50	Urea Nitrogen	10	mg/dL	Urea Nitrogen	10	mg/dL	Urea Nitrogen	10	mg/dL	Urea Nitrogen
12	101	65	M	2010-01-01	10:55	Creatinine	1.2	mg/dL	Creatinine	1.2	mg/dL	Creatinine	1.2	mg/dL	Creatinine
13	101	65	M	2010-01-01	11:00	Glucose	100	mg/dL	Glucose	100	mg/dL	Glucose	100	mg/dL	Glucose
14	101	65	M	2010-01-01	11:05	Calcium	10	mg/dL	Calcium	10	mg/dL	Calcium	10	mg/dL	Calcium
15	101	65	M	2010-01-01	11:10	Sodium	140	mg/dL	Sodium	140	mg/dL	Sodium	140	mg/dL	Sodium
16	101	65	M	2010-01-01	11:15	Potassium	4.0	mg/dL	Potassium	4.0	mg/dL	Potassium	4.0	mg/dL	Potassium
17	101	65	M	2010-01-01	11:20	Magnesium	2.0	mg/dL	Magnesium	2.0	mg/dL	Magnesium	2.0	mg/dL	Magnesium
18	101	65	M	2010-01-01	11:25	Phosphorus	3.0	mg/dL	Phosphorus	3.0	mg/dL	Phosphorus	3.0	mg/dL	Phosphorus
19	101	65	M	2010-01-01	11:30	Bilirubin	1.0	mg/dL	Bilirubin	1.0	mg/dL	Bilirubin	1.0	mg/dL	Bilirubin
20	101	65	M	2010-01-01	11:35	Albumin	4.0	g/dL	Albumin	4.0	g/dL	Albumin	4.0	g/dL	Albumin
21	101	65	M	2010-01-01	11:40	Triglycerides	150	mg/dL	Triglycerides	150	mg/dL	Triglycerides	150	mg/dL	Triglycerides
22	101	65	M	2010-01-01	11:45	HDL Cholesterol	50	mg/dL	HDL Cholesterol	50	mg/dL	HDL Cholesterol	50	mg/dL	HDL Cholesterol
23	101	65	M	2010-01-01	11:50	LDL Cholesterol	150	mg/dL	LDL Cholesterol	150	mg/dL	LDL Cholesterol	150	mg/dL	LDL Cholesterol
24	101	65	M	2010-01-01	11:55	Total Cholesterol	200	mg/dL	Total Cholesterol	200	mg/dL	Total Cholesterol	200	mg/dL	Total Cholesterol
25	101	65	M	2010-01-01	12:00	Urea Nitrogen	10	mg/dL	Urea Nitrogen	10	mg/dL	Urea Nitrogen	10	mg/dL	Urea Nitrogen
26	101	65	M	2010-01-01	12:05	Creatinine	1.2	mg/dL	Creatinine	1.2	mg/dL	Creatinine	1.2	mg/dL	Creatinine
27	101	65	M	2010-01-01	12:10	Glucose	100	mg/dL	Glucose	100	mg/dL	Glucose	100	mg/dL	Glucose
28	101	65	M	2010-01-01	12:15	Calcium	10	mg/dL	Calcium	10	mg/dL	Calcium	10	mg/dL	Calcium
29	101	65	M	2010-01-01	12:20	Sodium	140	mg/dL	Sodium	140	mg/dL	Sodium	140	mg/dL	Sodium
30	101	65	M	2010-01-01	12:25	Potassium	4.0	mg/dL	Potassium	4.0	mg/dL	Potassium	4.0	mg/dL	Potassium
31	101	65	M	2010-01-01	12:30	Magnesium	2.0	mg/dL	Magnesium	2.0	mg/dL	Magnesium	2.0	mg/dL	Magnesium
32	101	65	M	2010-01-01	12:35	Phosphorus	3.0	mg/dL	Phosphorus	3.0	mg/dL	Phosphorus	3.0	mg/dL	Phosphorus
33	101	65	M	2010-01-01	12:40	Bilirubin	1.0	mg/dL	Bilirubin	1.0	mg/dL	Bilirubin	1.0	mg/dL	Bilirubin
34	101	65	M	2010-01-01	12:45	Albumin	4.0	g/dL	Albumin	4.0	g/dL	Albumin	4.0	g/dL	Albumin
35	101	65	M	2010-01-01	12:50	Triglycerides	150	mg/dL	Triglycerides	150	mg/dL	Triglycerides	150	mg/dL	Triglycerides
36	101	65	M	2010-01-01	12:55	HDL Cholesterol	50	mg/dL	HDL Cholesterol	50	mg/dL	HDL Cholesterol	50	mg/dL	HDL Cholesterol
37	101	65	M	2010-01-01	13:00	LDL Cholesterol	150	mg/dL	LDL Cholesterol	150	mg/dL	LDL Cholesterol	150	mg/dL	LDL Cholesterol
38	101	65	M	2010-01-01	13:05	Total Cholesterol	200	mg/dL	Total Cholesterol	200	mg/dL	Total Cholesterol	200	mg/dL	Total Cholesterol
39	101	65	M	2010-01-01	13:10	Urea Nitrogen	10	mg/dL	Urea Nitrogen	10	mg/dL	Urea Nitrogen	10	mg/dL	Urea Nitrogen
40	101	65	M	2010-01-01	13:15	Creatinine	1.2	mg/dL	Creatinine	1.2	mg/dL	Creatinine	1.2	mg/dL	Creatinine
41	101	65	M	2010-01-01	13:20	Glucose	100	mg/dL	Glucose	100	mg/dL	Glucose	100	mg/dL	Glucose
42	101	65	M	2010-01-01	13:25	Calcium	10	mg/dL	Calcium	10	mg/dL	Calcium	10	mg/dL	Calcium
43	101	65	M	2010-01-01	13:30	Sodium	140	mg/dL	Sodium	140	mg/dL	Sodium	140	mg/dL	Sodium
44	101	65	M	2010-01-01	13:35	Potassium	4.0	mg/dL	Potassium	4.0	mg/dL	Potassium	4.0	mg/dL	Potassium
45	101	65	M	2010-01-01	13:40	Magnesium	2.0	mg/dL	Magnesium	2.0	mg/dL	Magnesium	2.0	mg/dL	Magnesium
46	101	65	M	2010-01-01	13:45	Phosphorus	3.0	mg/dL	Phosphorus	3.0	mg/dL	Phosphorus	3.0	mg/dL	Phosphorus
47	101	65	M	2010-01-01	13:50	Bilirubin	1.0	mg/dL	Bilirubin	1.0	mg/dL	Bilirubin	1.0	mg/dL	Bilirubin
48	101	65	M	2010-01-01	13:55	Albumin	4.0	g/dL	Albumin	4.0	g/dL	Albumin	4.0	g/dL	Albumin
49	101	65	M	2010-01-01	14:00	Triglycerides	150	mg/dL	Triglycerides	150	mg/dL	Triglycerides	150	mg/dL	Triglycerides
50	101	65	M	2010-01-01	14:05	HDL Cholesterol	50	mg/dL	HDL Cholesterol	50	mg/dL	HDL Cholesterol	50	mg/dL	HDL Cholesterol
51	101	65	M	2010-01-01	14:10	LDL Cholesterol	150	mg/dL	LDL Cholesterol	150	mg/dL	LDL Cholesterol	150	mg/dL	LDL Cholesterol
52	101	65	M	2010-01-01	14:15	Total Cholesterol	200	mg/dL	Total Cholesterol	200	mg/dL	Total Cholesterol	200	mg/dL	Total Cholesterol
53	101	65	M	2010-01-01	14:20	Urea Nitrogen	10	mg/dL	Urea Nitrogen	10	mg/dL	Urea Nitrogen	10	mg/dL	Urea Nitrogen
54	101	65	M	2010-01-01	14:25	Creatinine	1.2	mg/dL	Creatinine	1.2	mg/dL	Creatinine	1.2	mg/dL	Creatinine
55	101	65	M	2010-01-01	14:30	Glucose	100	mg/dL	Glucose	100	mg/dL	Glucose	100	mg/dL	Glucose
56	101	65	M	2010-01-01	14:35	Calcium	10	mg/dL	Calcium	10	mg/dL	Calcium	10	mg/dL	Calcium
57	101	65	M	2010-01-01	14:40	Sodium	140	mg/dL	Sodium	140	mg/dL	Sodium	140	mg/dL	Sodium
58	101	65	M	2010-01-01	14:45	Potassium	4.0	mg/dL	Potassium	4.0	mg/dL	Potassium	4.0	mg/dL	Potassium
59	101	65	M	2010-01-01	14:50	Magnesium	2.0	mg/dL	Magnesium	2.0	mg/dL	Magnesium	2.0	mg/dL	Magnesium
60	101	65	M	2010-01-01	14:55	Phosphorus	3.0	mg/dL	Phosphorus	3.0	mg/dL	Phosphorus	3.0	mg/dL	Phosphorus
61	101	65	M	2010-01-01	15:00	Bilirubin	1.0	mg/dL	Bilirubin	1.0	mg/dL	Bilirubin	1.0	mg/dL	Bilirubin
62	101	65	M	2010-01-01	15:05	Albumin	4.0	g/dL	Albumin	4.0	g/dL	Albumin	4.0	g/dL	Albumin
63	101	65	M	2010-01-01	15:10	Triglycerides	150	mg/dL	Triglycerides	150	mg/dL	Triglycerides	150	mg/dL	Triglycerides
64	101	65	M	2010-01-01	15:15	HDL Cholesterol	50	mg/dL	HDL Cholesterol	50	mg/dL	HDL Cholesterol	50	mg/dL	HDL Cholesterol
65	101	65	M	2010-01-01	15:20	LDL Cholesterol	150	mg/dL	LDL Cholesterol	150	mg/dL	LDL Cholesterol	150	mg/dL	LDL Cholesterol
66	101	65	M	2010-01-01	15:25	Total Cholesterol	200	mg/dL	Total Cholesterol	200	mg/dL	Total Cholesterol	200	mg/dL	Total Cholesterol
67	101	65	M	2010-01-01	15:30	Urea Nitrogen	10	mg/dL	Urea Nitrogen	10	mg/dL	Urea Nitrogen	10	mg/dL	Urea Nitrogen
68	101	65	M	2010-01-01	15:35	Creatinine	1.2	mg/dL	Creatinine	1.2	mg/dL	Creatinine	1.2	mg/dL	Creatinine
69	101	65	M	2010-01-01	15:40	Glucose	100	mg/dL	Glucose	100	mg/dL	Glucose	100	mg/dL	Glucose
70	101	65	M	2010-01-01	15:45	Calcium	10	mg/dL	Calcium	10	mg/dL	Calcium	10	mg/dL	Calcium
71	101	65	M	2010-01-01	15:50	Sodium	140	mg/dL	Sodium	140	mg/dL	Sodium	140	mg/dL	Sodium
72	101	65	M	2010-01-01	15:55	Potassium	4.0	mg/dL	Potassium	4.0	mg/dL	Potassium	4.0	mg/dL	Potassium
73	101	65	M	2010-01-01	16:00	Magnesium	2.0	mg/dL	Magnesium	2.0	mg/dL	Magnesium	2.0	mg/dL	Magnesium
74	101	65	M	2010-01-01	16:05	Phosphorus	3.0	mg/dL	Phosphorus	3.0	mg/dL	Phosphorus	3.0	mg/dL	Phosphorus
75	101	65	M	2010-01-01	16:10	Bilirubin	1.0	mg/dL	Bilirubin	1.0	mg/dL	Bilirubin	1.0	mg/dL	Bilirubin
76	101	65	M	2010-01-01	16:15	Albumin	4.0	g/dL	Albumin	4.0	g/dL	Albumin	4.0	g/dL	Albumin
77	101	65	M	2010-01-01	16:20	Triglycerides	150	mg/dL	Triglycerides	150	mg/dL	Triglycerides	150	mg/dL	Triglycerides
78	101	65	M	2010-01-01	16:25	HDL Cholesterol	50	mg/dL	HDL Cholesterol	50	mg/dL	HDL Cholesterol	50	mg/dL	HDL Cholesterol
79	101	65	M	2010-01-01	16:30	LDL Cholesterol	150	mg/dL	LDL Cholesterol	150	mg/dL	LDL Cholesterol	150	mg/dL	LDL Cholesterol
80	101	65	M	2010-01-01	16:35	Total Cholesterol	200	mg/dL	Total Cholesterol	200	mg/dL	Total Cholesterol	200	mg/dL	Total Cholesterol
81	101	65	M	2010-01-01	16:40	Urea Nitrogen	10	mg/dL	Urea Nitrogen	10	mg/dL	Urea Nitrogen	10	mg/dL	Urea Nitrogen
82	101	65	M	2010-01-01	16:45	Creatinine	1.2	mg/dL	Creatinine	1.2	mg/dL	Creatinine	1.2	mg/dL	Creatinine
83	101	65	M	2010-01-01	16:50	Glucose	100	mg/dL	Glucose	100	mg/dL	Glucose	100	mg/dL	Glucose
84	101	65	M	2010-01-01	16:55	Calcium	10	mg/dL	Calcium	10	mg/dL	Calcium	10	mg/dL	Calcium
85	101	65	M	2010-01-01	17:00	Sodium	140	mg/dL	Sodium	140	mg/dL	Sodium	140	mg/dL	Sodium
86	101	65	M	2010-01-01	17:05	Potassium	4.0	mg/dL	Potassium	4.0	mg/dL	Potassium	4.0	mg/dL	Potassium
87	101	65	M	2010-01-01	17:10	Magnesium	2.0	mg/dL	Magnesium	2.0	mg/dL	Magnesium	2.0	mg/dL	Magnesium
88	101	65	M	2010-01-01	17:15	Phosphorus	3.0	mg/dL	Phosphorus	3.0	mg/dL	Phosphorus	3.0	mg/dL	Phosphorus
89	101	65	M	2010-01-01	17:20	Bilirubin	1.0	mg/dL	Bilirubin						

# Detecting errors in Cats and Dogs

Intentionally introduced up to 50% errors in cat image labels, and up to 30% error for both cats and dogs

UDC automatically identified incorrectly labelled images

AI accuracy improvement after UDC was between **20% to 30%**



# Detecting errors in Vehicles (four labels)

Intentionally introduced up to 70% error across all four types of vehicle images

UDC automatically identified incorrectly labelled images.

AI accuracy improvement after UDC was up to **45%**



# Detecting errors in embryo images

PGT-A is an invasive technique where an embryo is biopsied and genetically tested for abnormalities like down syndrome

Nature of PGT-A testing can result in “errors”: overestimation of embryos being genetically abnormal and thus unnecessarily discarded

UDC identified 37% of embryo images that were likely mis-labelled

AI accuracy improvement after UDC was up to 10%.

## Detect Genetic Integrity





# Beyond errors: detecting difficult to assess X-Ray images

UDC identified 'noisy' images that were difficult for both expert radiologist and AI algorithms to identify pneumonia

Noisy images had a negative impact on AI training, accuracy and generalizability

AI accuracy improvement after UDC was up to 25%

Useful as a triage tool to identify scans that may be difficult to diagnose and may need additional tests

**Detect pneumonia**





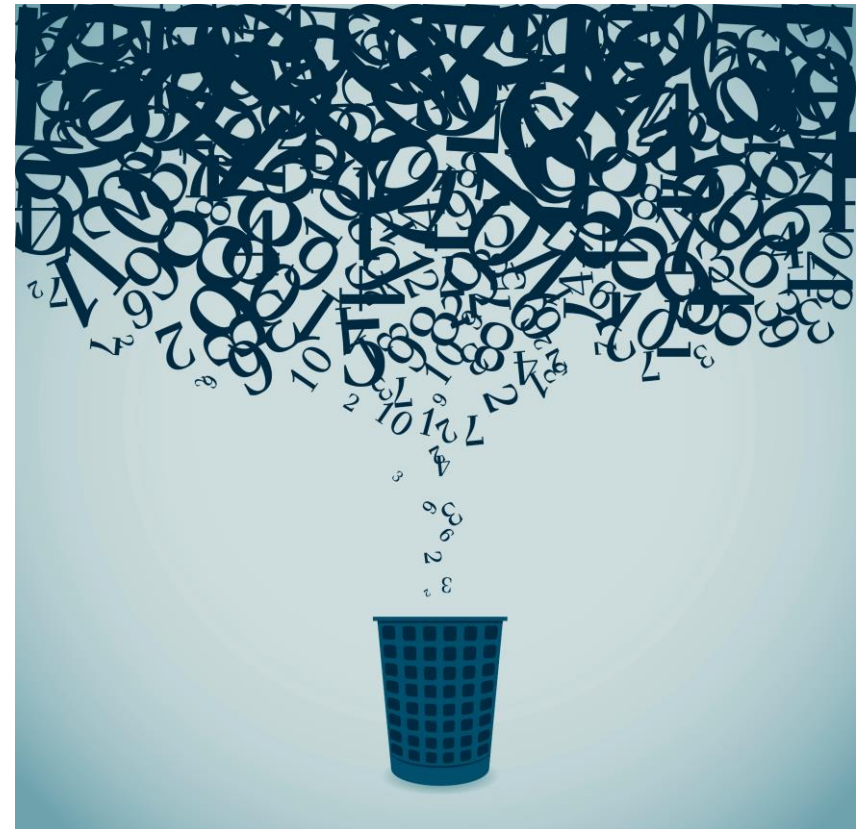
# Other potential applications of the UDC?

Error detection of private data that can't be manually verified

Scalable and practical error detection of massive datasets

Detection of anomalies in data

Labeling or prediction of difficult to detect or predict data





# Presagen

A.I. Enhanced Healthcare

Dr Don Perugini  
Co-Founder

e: Don@Presagen.com  
l: <https://linkedin.com/in/donperugini>  
t: @DrDonPerugini  
w: Presagen.com

[info@Presagen.com](mailto:info@Presagen.com)  
[www.Presagen.com](http://www.Presagen.com)  
[www.linkedin.com/company/Presagen/](https://www.linkedin.com/company/Presagen/)  
<https://twitter.com/Presagen>