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

I: https://linkedin.com/in/donperugini

t: @DrDonPerugini

w: Presagen.com

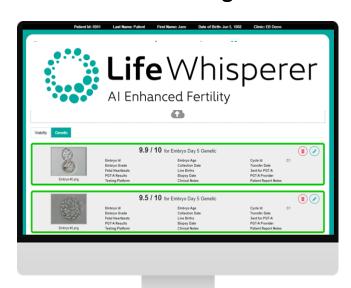


Medical data is inherently poor quality



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

Artificial Intelligence



Precision Health

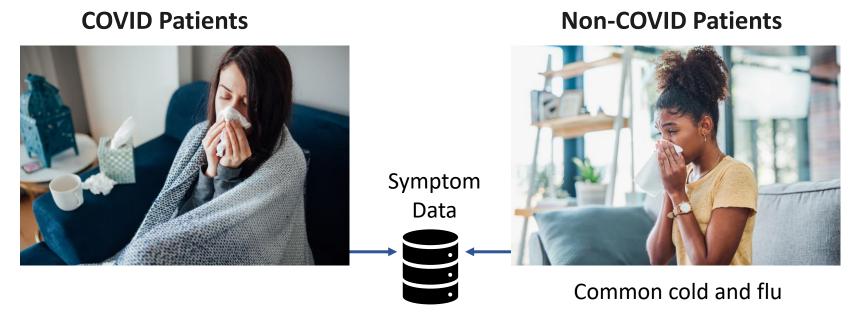


Clinical trials



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

ed with COVID patients?



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

COVID Patients



Symptom Data







Common cold and flu

RAT used to diagnose COVID



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





<3% error in detecting Non-COVID



Symptom

Data



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

Artificial Intelligence can be used to automatically and reliably detect inherent errors in medical data

Al algorithm called UDC

Identifies inconsistencies or errors in data

Preserves privacy: no manual verification needed

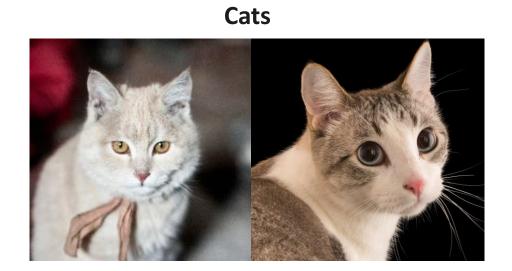
Turning AI onto itself...



Patent: PCT/AU2021/000028

Consider AI trained to detect cats and dogs

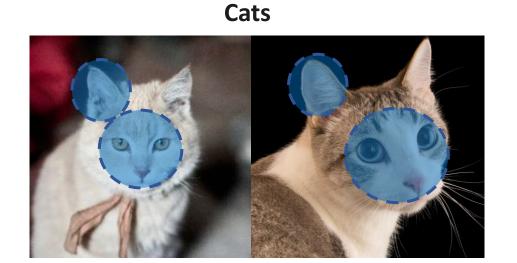


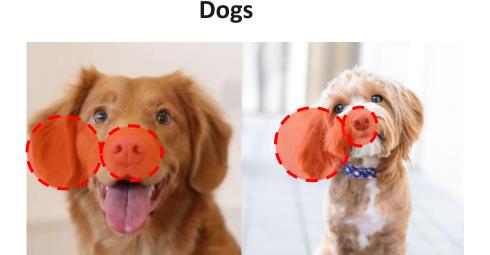






Al learns patterns and features from historical data to distinguish between cats and dogs





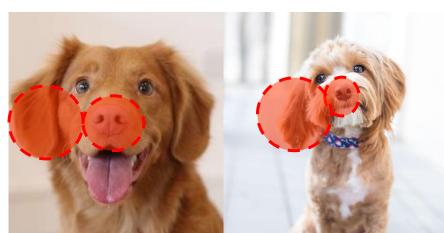
Data needs to be labelled as 'Cat' or 'Dog' to train the Al

Errors in data creates inconsistencies during Al training

Cats **Dogs**



Al can't match the features that look different

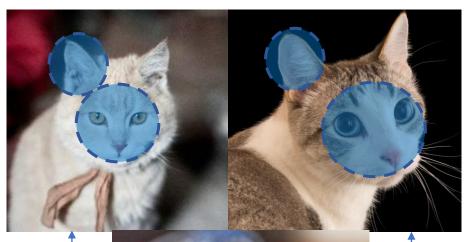


More often during training the Al will match the features associated with "Dogs" and think it is incorrect

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

'Untrainable' consistently incorrect data are likely errors

Cats **Dogs**



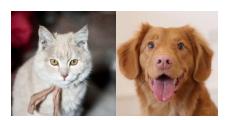
Al can't match the features that look different

More often during training the Al 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



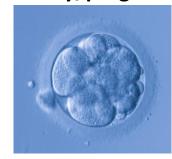
X-Ray images to detect pneumonia



Vehicles



Embryo images to predict viability/pregnancy



Non-Image Record Medical Data

2	Johnson		This Peak The Peak (6) Th	mounts	300,746	Or prost	Ser.	Asia Pros. Ami Pros. (611)
that across the	Assumb	349.00	1.007003-01-0079	OLIV UFS	10		DMR	8-340,PM, LT 25,4679
tale Nov. No.	AMUFE	4.0 (6)	1,000,7159,859	SHRUFA	4	N	10 H-8	120346-1131-859
Chraniphose Rain	ARLUFA.	3.95 66	LENGTH IN WITH	GHR UFS	N	N.	U.H.W	0.28602-U.SL-W79-
BLADN'ON No.	APPLIES.	XM III	A DECISION IN THE ARTHR.	MARLES		N	11114	A HOUSE AT THE METER
Julian Clance prices flar	ARTUFE	2.81.66	1,000,051.64.8679	GLM UPS	N	N .	11,16.00	0.7503 M-01.06 ARTH
Core with Facility	ARDFI	199.66	1,100 km² ti ne 4679	DEST-UPS		4	12.01-00	3-2769T LL LL ACTV
Single of the Greatest Asso	APILLER	214.6	TRIALTS Was	-DAM UPS	N	N	ALM M	5.340605-35.71,4679
Lorder's drawle drop flaci	SERFLAPE.	134.65	FTMOSETIA MAIN	GENUTY	3	N	123148	3 (00000 1L 00 da70)
Opinior scientists flux:	ANUFL	1,41.00	1,0907376979	MALLEY		h	418.4	4.29403-U-41-WV
saftering below, float	ARIOPS	180	0.1603653-30-4619	MANUFE	9	7	44.00	9,079035-11,05-4079
Comments for	ARUFT	1170 0	1,154,053,74-8079	OR PRIVATE		4	45.16.00	4,0440,0150,000
Sty contrasts seems flat	ANIMA	Lua	1300713-879	SHIP UPS	h	N	444	\$1050H-1030-809
Charlosols plants flux	SHILLER.	344	LINESCED MINE	MANAGES	8	N.	45 (4.00	1,00000 (1.10 (679)
The sect of security Rights entire, but	4.000	104	FIGHTED MAN	MAUFE	N	A	403146	3.5000 4.90 db7F
carbo toron - 100 attent - Nat	AR WES	HIM	1.00001 m m/m	MANUF	h	h	11.15-6	4.3766 4 ft-859
COST - school containers - (HE) various has	4 SHAPE	337.00	\$20000 France	GERLUFT.	10	N	0.00	130700 140 4079
Chairman - Committee Staff Disc.	ABUFF	N.H.W.	LINDRESS ATT	CHRUFF	A	A	拉矿商	4,0000 U.B.#76
regisatio - venosifititi- his.	0.6LUPS	150	1.180000.1 or 4879	GER LIFE		19	43.94.66	6.07530-01.36-009-
Contractor feet	D.M.LOP'S	Table 1	1,0040347-009	MAUPE	3	h	1241 M	6.34675 63.93 4679
Liddela synthesis files	ABUPL		1,009/61/1/ (8/4)	DENTUFF		N	拉杯鄉	6.8896.4TL#09
Death facility Res	SHIDE	5.00.00	1.164000 3 Strategy	OK NOW PE		in .	4211.00	0.000002 55 98 9879
etorical routio, flac	ABUFE	A 13 at	1.11(000) (1" (070)	MHUFF	3	N	411.0	CHRIST HALLET
Leader's more flat	BUTUFF	A39 (6)	6.6940C370-657P	MINUTE	N	9	HHA	\$100000 11 to Miles
Decale for	# SHAPE	200	1.36740E1 IN 6878	MANUFA		*	11-0-0	5.30400 5.78 of TO
Crimingson Science San.	CHUFF	10.0	3,100033,75.4007	ANNUFS	h .	N	25 47 48	8,98095 & 13 METP
hing to whose fac	ADVINE	0.50 46	LINGALLTWIN	OHNUFS.	4	N .	信用曲	0.007501-00.04-0079
House, s. a Phase On Earth Ass.	BRUFE	101.0	1.100500137 WEST	CHRUFT	h	4	41.0 @	6.00000 6.00 4070
Maryon gareen Rape	AMMER	SHA	LANCYLL REPT	MINUTE	A	N	504	4 (274KD 11 (0.4674)
HENTE Engage for	ARUFE	3.29 (6)	1040014-819	deal of a		N .	10.04 (8)	\$1000 U.S ATF
conferred religion float	ANUFS		LT0961 (KW19	2630 UPI		N .	出現者	5.064602-51-52 off?#
Nati	ATOURS	10.0	1,040,43 (7,4619	AND LIFE	N	16	40.11 46	5 10 7005 35 31 4679

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

Al 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.

Al 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

Al 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

Al 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



Works on any data...

UDC applied to non-image record-based medical data

UDC identified errors that were intentionally introduced

UDC can be applied non-AI healthcare use cases, e.g. errors in clinical trial data, or other data used in the treatment of patients

Non-Image Record Medical Data

Fig	Lindony Raspe	This Peak Tile Peak (877)	Pathershick	World , Strape	Oa person	Sarr	Asse Pres. Age; Pres. 4815
IS that across the	A STAPE YOUR	5.29790813 4K-48739	UNITED AFTS	in .		-33 M dl	LONG U.S. APP
int nise New Nex	AMMERICAN ASSESSED.	LOTELY CO. MAN	OWNER	4	No.	10 99-88	\$26046-LC314879
CE True phone field	WHEN THE PARTY AND PERSONS NAMED IN	A STREET, NAME AND POST OF	GLES WES	N		431,74 (46)	0.296021-03.00 48795
DE SELECTION PRO	WATER THE REAL PROPERTY.	A designation has placed by	ARREST LAWY	4	N	10115-0	A STOCKET ALL TO MATER
III. when chance project flow	WATEHAN THEM	1,100305-0-M-3870	GRAN NUMBER	16	N.	145,94 (8)	0.090304-03-98-98094
M. Core-with Floridge	SHOW THE THEFT	L 100ke/ u be with	GROW LOPE	10	-	150,001/00	0.0750HT-13.53.467H
EF Einge of the Greaters Au-	APPLIES SHOW	A SAFEKET LAT MAKE	-DEMONSTR	N	N	A0.00 val	5.340605-35.71 (679)
Of upday a drawle one had	CONTRACT COURSE	A LINCOLL ST MANY	SERVICE	34	N	12:11:40	3 (80000 11 or 40°0")
Of glocial sciences has	SHAPE SHIP	5.50 KDC 3.70 METE	000 00 LOW'S	9	14	45.00-46	6.29403-03-63 WE'VE
(a) colony bean feet	ARREST LINE	0.0003003-30-0010	ARREST AND S	9	7	14.75.46	9 (F190) \$ 11.05 (679)
11, comments for	ARREST MARK	1.15400EQ 24-9ETF	-04 04 LAPE	74	19	140,741.00	4 (866) 00 10 4019
SE sky construint present flag	AND DESCRIPTION OF THE PERSON NAMED IN COLUMN	S SHART'S ON WITH	SHOP LAFE	h	74	10.00	0.075099-14.00-98799
S Extractoración prometa filoso	CHARLETT THE STATE	S. ANTONOMY SHE SHEET	ARREST AND ADDRESS.	16	10	A40 04 188	9-2400MG 53-30 467W
14. The world secure Highline entires had	1806 108	LICHTED DWW.	HARRIST LANDS	16	in.	40/31 46	3.0000 KH 607
Shorter torons - 1922 without - Park	ARRIVES AND IN	A DESCRIPTION OF THE PARTY.	SHAPE SAFE	A	*	11.11-6	4.30 kin a fruitti
18, 1865 - school contains - 250 virging his	SHARE SOUR	5.20000 S 75.40707	GRIDENAT'S	16	N	10.4440	4,007000-040-0079-
17 Tel-1008 - General Re-Out- Sec.	ADDRESS NAME.	1,000,000,00,000	OF REAL PROPERTY.	N .	Pa .	HEAT ME	4.8000 11.6 400
id registric remonistic fee.	DESCRIPTION AND ADDRESS.	E EMPORTS CAN ARREST	GENERALFE		N .	12/16/46	6.277236-13.36-4679
18 continue fac	DESCRIPTION OF THE PARTY.	LANGERS OF MICH.	MANAGER	31	N.	104116	A Joseph St. Market
(5-ratiota synthesis Nee	ASSESSED AND MARKET	LINEAU PLANT	DESCRIPTION OF THE PERSON NAMED IN		A	155,94100	5.3659C-6-T1-6659C
23 Janes Super State	SHIDTS SHIP	1. Telephone 2 (tol page 19)	0630 MF6		No.	12014	0.000002 25 98 9879
15 especial resists fluir	ARRESTS ASSESSED.	A RESIDENCE OF MATER	00.001,075	31	N.	40.00	CARROLL STATE SETT
25-color's moss flat	SHITTAPE ASSESS.	6.00000000 70 WAR	00 00 LAFE	N	74	12.15.00	COMMENTAL STATE OF THE PARTY.
D4.Decople-Nati	WHITE REPORT	1.26240011 IN 1879	AND RELATED		76	143-99-08	5.300,600 to 50 sector
of decorator force file.	CHARLES SHOW	3,0000033,75-4000	-back sumb	10	N	Assistant	0.500,005 to 73 sector
Shows to winese his:	-5-29-23-PN -0-00-WE	EDWINALLT MEDI	OWNER		No.	40 h 46	IN CONTROL OF THE WATER
27 Property is a Phase Dr. Earth Ave.	BRIDE SHEET	A LIMITAGE AND WATER	GROWINGS.		in .	11.40 (8)	6.000MG-6-95-467F
16 future gaves hav-	SALAMAN NAME OF	ELMONTH LINE METER	000 M 144FW	.04	N	12.00	4 CTM60 13 01 4674
DELEVEL Exception for	4 R UFS 3-26 (6)	2.004/D018-01-08709	debt sups	76	No.	120.04 (8)	0.00400, 42.104 decree
Minerly tric radigors facilities	SHOPS SHOW	A DOMESTIC MAN	08/30 NUFE		14	10.00	9.06w862-24-50 s8709
interest	ATTOMY SHOW	5.3840.4G.17 MTT	-09-95 to FT	16	N.	44 Ft all	\$10,000 10.31 0070

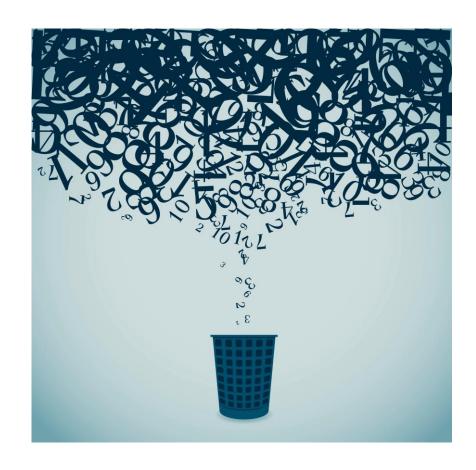
Other potential applications of the UDC?

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





Dr Don Perugini Co-Founder

e: Don@Presagen.com

I: https://linkedin.com/in/donperugini

t: @DrDonPerugini

w: Presagen.com

info@Presagen.com www.Presagen.com

www.linkedin.com/company/Presagen/

https://twitter.com/Presagen_