# AUTOMATED INTERPRETATION THE OTHER, BETTER, WISER AI

- Machine Learning models often claim to predict rules by analysing data and answers.
- This is empirical analysis to an extreme, but risks omitting extraneous data created from unpredictable factors like human behaviour.

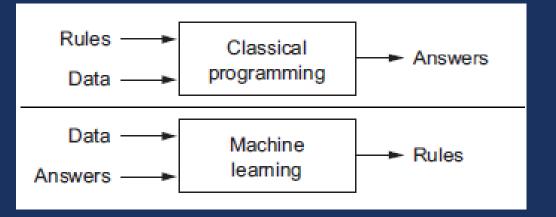


Figure I - A comparative model of procedural programming and machine learning.

observing seminars as case • By studies, and evaluating them as qualitative data, it was possible to gain insight into socio-industrial trends and methods through the perspective of workers within that context.

• These seminars can offer a large set of rich qualitative data, by looking at data taken from individual seminars and looking at the meta data of the collective set.

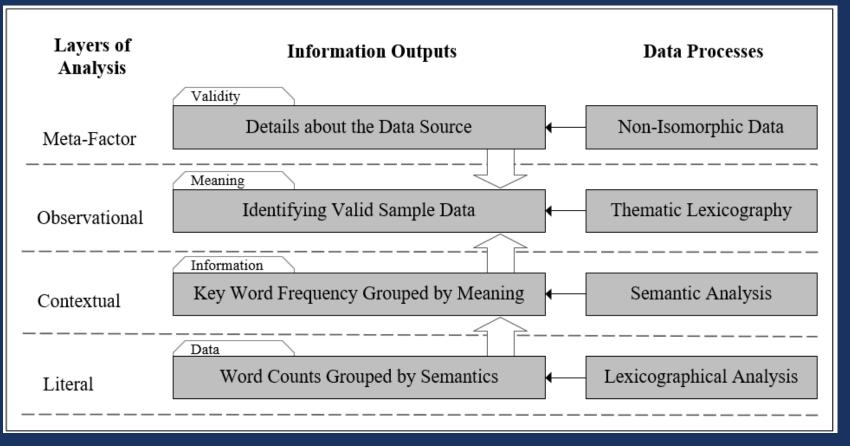


Figure 2 – A systematic, multi-modal model of interpretive content analysis for application of deep learning for complex qualitative data sets.

# **Problem Analysis and Formative Challenges**

**Empirical Science** Interpretive Science Automated Interpretation **Grounded Theory** Machine Learning

Qualitative Analysis

**Content Analysis** 

### A. Limitations of Automated Qualitative Analysis

- Expensive to initially gather data
- Processing and analysis is highly subjective
- Hard to replicate, hard to control

/erification

Validation

## B. Limitations of Procedural Quantitative Analysis

- Abstract and prone to omit non-isomorphic data
- Predisposed to unobserved biases and
- Can lacks semantic and contextual understanding

# **New Methods of Content Analysis and Experimental Findings**

#### Literal Data Modelling

Lexicographical analysis and linear regression are applied to highlevel data from the sample (such as seminar titles) and used to identify localised patterns within the data set.

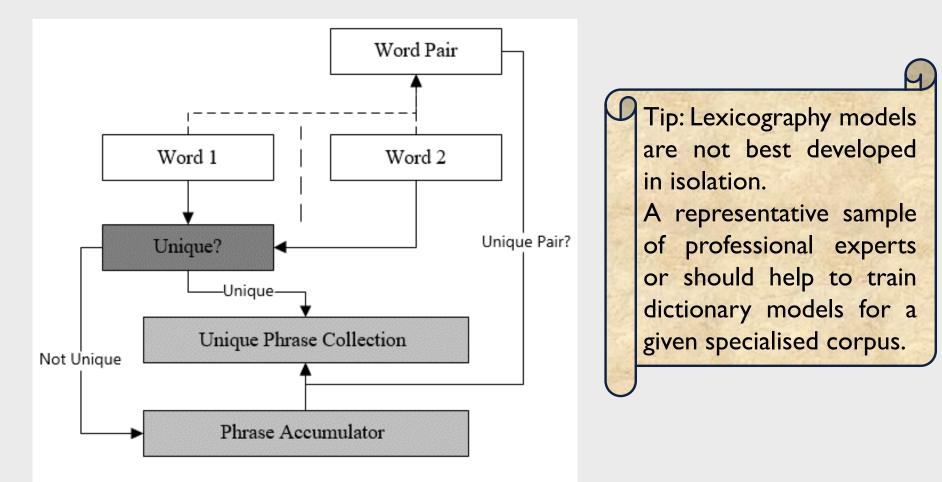


Figure 3 – A word-pairing algorithm combined with natural language processing is used to model key trends and factors at a high-level from the data set.

#### **Observational Data Modelling** 2.

Word pairs are semantically grouped and modelled to find aggregate trends across the samples using recursive analysis.

		2017					2018						2019		
Ch		Rank	- F	Concat			Rank /		Concat	Ch			Rank	- F	Concat
-12	•	3	19	Animation		•	1	23	Reflection		-2	•	1	14	Your
-7	•	3	19	Indie		•	2	15	Player		0	$\leftrightarrow$	1	14	Reflection
-7	•	9	8	Generation		۲	3	14	Your		-3	•	2	9	Art
-7	•	9	8	Procedural		•	4	13	Story		-3	•	3	7	Creating
-6	۸	10	7	Day		•	5	12	Making		-4	•	3	7	Building
-6	•	10	7	No		۲	5	12	Art		-7	•	3	7	Bootcamp
-6	۸	10	7	UX		۲	6	11	Creating		-1	•	4	6	Making
-5	۸	2	21	Your		۲	6	11	Animation		-9	•	4	6	Play
-5	۸	8	10	Character		۲	7	10	Level		-5	۸	4	6	World
-5	۸	11	6	Direction		۸	7	10	Day		2	۷	4	6	Player
-5	۸	11	6	Our		۲	7	10	Al		-3	•	4	6	Al
-5	۸	11	6	Programmer		۲	7	10	Building		-9	•	5	5	Your Game
-5	۸	11	6	Social		۸	7	10	Development		-2	۸	5	5	Level

Figure 4 – A recursive model is used to identify initial trends over aggregate data.

Tip: Compile a lexicography filtering out stop words and then compile one which retains them, analysing both.

Stop words lack semantic meaning when analysed in isolation, but add details to word pairs. E.g. 'Your game' is prescriptive, but 'our game' is reflective.

Meta-factor Data Source Modelling 4.

### 3. **Contextual Language Modelling**

Deep learning is applied to the full content of each qualitative data set to model key topics to Sum Frequency of 15 Most Frequent Words Per form a corpus-based language Category With Semantic Value model.

Tip: Dictation programs were found to be up to 37% more accurate than auto- captioning libraries.

'Big Data' on the cloud has allowed language process of timbre, accents, and intonation from speaker(s) to improve analysis and recognition.

350 -313 300 245 S 250 231 Frequency 500 Fr 182 105 59 50 Methods User Platform Commercial Artistic ools

Figure 5 - A corpus-based semantic model is trained to recognise aggregated trends.

Observations data-mined from the data source offers rich details to facilitate multiple-linear regression to identify trends, while maintaining power of the experiment as the sample size

	Damaa	manhia				fun	than rad	~~~~~	ad and shrinks
Game Dev	Demog		Ter	line	— IS	IUI	uner reg	21 ess	ed and shrinks.
Count Phrase		Research Count Phrase		Tooling Count Phrase			C	•	
Tool	12 Proc	Too							
30 Bug	12 FIOC Too	44 Rig						A	
55 565	100		Studio Size						
25 Player	Solo/Duo	Small		AAA					
18 Designer	Count Phrase		Count Phrase		e Count Phra		ase		Tip:There is always
18 Build	Tool		Tool		Tool				
17 Simple	16 Flow		27 Bug	44 Die			_ 8	non-isomorphic	
14 Editor				bject			Total	. 5	A TO DE TRANSPORTE AND A TO DE TRANSPORTE A
14 Test	15 Editor		Tech Art		Tools			_ [	data which can't be
14 QA 13 Artist	13 Proc-Gen	Count	Phrase	Count	Phrase	Count	Phrase	_	A REAL REAL PROPERTY AND A REAL PROPERTY A REAL PROPERTY AND A REA
13 Script	10 Engine		Tool	22	Tool Bug	129 44		- E	anticipated or
12 Data	10 Visual 10 Data	44	KIg	52	bug	44	RIE		
12 Trigger	9 Productivit	29	Animation	31	Editor	38 Data 33 Editor			which will be
11 Fix	9 Level-Desi		Kinematic		Player				
44 Dester	8 Menu 16 Art 8 Player 14 Ch 8 Strategy 13 Pe		Animator	25	Data	33 Animation			tainted by observed
			Artist	ist 21		32	Player		
			Character	21	Build	32 Bug			or unobserved
			Performance		QA		Artist		
	7 Zone	Iteration				Build	- 8	biases.	
·					Simple		Scripting		Sector March String Contractor
			Scripting		Test Iteration	29	Iteration	- 2	to contraction and the
			Rigging Tech-Art		Flow		ua Animator		
			Pipeline		Artist		Designer		
						24	Viennesie		

Figure 6 – The corpus-based semantic model is cross-referenced against meta-data from the source to find deeper trends from the aggregate data.

### **RESEARCHER: MR BRIAN WILLIAM PACKER SUPERVISOR: PROFESSOR SIMEON KEATES**

