

Supplementary Information

Supplementary Table 1: Number of concepts and sentences per batch

No. of reports	No. of concepts	No. of sentences
Batch 1 (n = 76 real)	970	818
Batch 2 (n = 15 real)	452	344
Batch 3 (n = 97 real)	2644	1960
Batch 4 (n = 76 synthetic)	970	818

Reports per year (excluding batch 4): 2019 (n = 4), 2020(n = 115), 2021 (n = 42), 2022 (n = 27).

Supplementary Table 2: Percentage of reports from Batch 1 containing each concept/human factor for each ethnic group calculated by using the formula shown at the bottom of this table

Concept	Asian	Black	DNR	MB	WO	WB	Average	SD
Organisation-Communication factor	83.33	100.00	75.00	100.00	66.67	75.00	83.33	13.94
External Environment-COVID	66.67	57.14	25.00	100.00	66.67	80.77	66.04	25.04
Organisation-Teamworking	83.33	71.43	50.00	100.00	33.33	57.69	65.96	23.99
Job/task-Assessment, investigation, testing, screening (e.g. holistic review)	66.67	42.86	50.00	100.00	50.00	63.46	62.16	20.60
Organisation-National guidance	16.67	71.43	50.00	100.00	16.67	71.15	54.32	33.22
Job/task-Care Planning	50.00	42.86	50.00	100.00	33.33	32.69	51.48	24.96
Person Patient-Physical characteristics	33.33	57.14	75.00	0.00	83.33	48.08	49.48	30.22
Technologies and Tools-issues	33.33	42.86	25.00	100.00	16.67	28.85	41.12	30.13
Job/task-Situation awareness (e.g. loss of helicopter view)	33.33	28.57	50.00	100.00	0.00	25.00	39.48	33.76
Person Staff-Slip or lapse (errors that tend to happen in routine tasks that people are doing without much conscious thought)	33.33	42.86	50.00	0.00	33.33	42.31	33.64	17.66
Organisation-Language barrier	50.00	0.00	0.00	100.00	16.67	0.00	27.78	40.37
Organisation-Documentation	50.00	14.29	25.00	0.00	16.67	53.85	26.63	21.21
Job/task-Monitoring	50.00	14.29	25.00	0.00	33.33	26.92	24.92	16.96
Organisation-Escalation/referral factor (including fresh eyes reviews)	16.67	28.57	50.00	0.00	16.67	36.54	24.74	17.52
Job/task-Obstetric review	33.33	14.29	25.00	0.00	33.33	34.62	23.43	13.83
Job/task-Dispensing, administering	0.00	0.00	100.00	0.00	0.00	13.46	18.91	40.09
Person Patient-Psychological characteristics (e.g. stress, mental health)	16.67	14.29	50.00	0.00	16.67	15.38	18.83	16.54
Internal Environment-Acuity (e.g. capacity of the maternity unit as a whole)	33.33	0.00	25.00	0.00	16.67	11.54	14.42	13.41
Person Staff-Training and education (e.g. attendance at antenatal classes)	16.67	0.00	50.00	0.00	0.00	15.38	13.68	19.45
Job/task-Risk assessment	33.33	28.57	0.00	0.00	0.00	13.46	12.56	15.24
Internal Environment-Physical layout and Environment	0.00	14.29	25.00	0.00	0.00	5.77	7.51	10.24
Person Staff-Decision error (errors in conscious judgments, decisions due to lack of knowledge and from misunderstanding of a situation)	0.00	0.00	0.00	0.00	16.67	23.08	6.62	10.46
Technologies and Tools (e.g. CTG, use of screening tools)-Interpretation (e.g. CTG)	0.00	14.29	0.00	0.00	0.00	21.15	5.91	9.40
External Environment-Geographical factor (e.g. Location of patient)	16.67	0.00	0.00	0.00	0.00	1.92	3.10	6.69

For example, the Organisation-Communication factor appeared in 7 out of 7 reports of the Black ethnic group, hence 100% occurrence. DNR: Data not received, WO: White Other, WB: White British, MB: Mixed background, SD: Standard Deviation.

Formula for computing the percentages: Number of reports in which a concept i appeared for ethnicity e /Total number of reports in ethnic group e .

Supplementary Table 3: Statistics about the investigation reports provided by HSIB (Batch 1)

Ethnic group	No. of reports	No. of concepts across the reports	Average no. of concepts per report
Asian	6	87	15
Black	7	81	12
Data not received	4	55	14
Mixed Background	1	13	13
White British	52	688	14
White Other	6	46	8
Total:	76	970	Average: 13

Supplementary Table 4: File and concept frequencies

Concept	Asian	Black	DNR	MB	WO	WB	Sum
Organisation-Communication factor	5(20)	7(12)	3(4)	1(4)	4(6)	39(93)	59(139)
Job/task-assessment, investigation, testing, screening (e.g. holistic review)	4(7)	3(5)	2(3)	1(1)	3(4)	33(73)	46(93)
External Environment-COVID	4(5)	4(8)	1(1)	1(1)	4(6)	42(71)	56(92)
Organisation-Teamworking	5(9)	5(11)	2(8)	1(1)	2(3)	30(58)	45(90)
Organisation-National guidance	1(2)	5(7)	2(4)	1(2)	1(1)	37(73)	47(89)
Person Patient-Physical characteristics	2(2)	4(6)	3(4)	—(—)	5(6)	25(41)	39(59)
Organisation-Documentation	3(3)	1(1)	1(1)	—(—)	1(1)	28(43)	34(49)
Person Staff-Slip or lapse (errors that tend to happen in routine tasks that people are doing without much conscious thought)	2(3)	3(4)	2(3)	—(—)	2(4)	22(30)	31(44)
Job/task-Care Planning	3(5)	3(6)	2(6)	1(1)	2(2)	17(20)	28(40)
Organisation-Escalation/referral factor (including fresh eyes reviews)	1(3)	2(4)	2(2)	—(—)	1(1)	19(28)	25(38)
Job/task-Obstetric review	2(2)	1(1)	1(1)	—(—)	2(3)	18(24)	24(31)
Job/task-Monitoring	3(5)	1(1)	1(1)	—(—)	2(2)	14(21)	21(30)
Technologies and Tools-issues	2(2)	3(5)	1(2)	1(1)	1(2)	15(17)	23(29)
Job/task-Situation awareness (e.g. loss of helicopter view)	2(4)	2(2)	2(2)	1(1)	0(—)	13(18)	20(27)
Person Staff-Decision error (errors in conscious judgments, decisions due to lack of knowledge and from misunderstanding of a situation)	—(—)	—(—)	—(—)	—(—)	1(2)	12(17)	13(19)
Technologies and Tools (e.g. CTG, use of screening tools)- Interpretation (e.g. CTG)	—(—)	1(3)	—(—)	—(—)	—(—)	11(13)	12(16)
Internal Environment-Acuity (e.g. capacity of the maternity unit as a whole)	2(3)	—(—)	1(2)	—(—)	1(1)	6(9)	10(15)
Person Patient-Psychological characteristics (e.g. stress, mental health)	1(1)	1(1)	2(3)	—(—)	1(1)	8(9)	13(15)
Job/task-Dispensing, administering	—(—)	—(—)	4(4)	—(—)	—(—)	7(9)	11(13)
Person Staff-Training and education (e.g. attendance at antenatal classes)	1(1)	—(—)	2(3)	—(—)	—(—)	8(9)	11(13)
Job/task-Risk assessment	2(2)	2(3)	—(—)	—(—)	—(—)	7(7)	11(12)
Organisation-Language barrier	3(7)	—(—)	—(—)	1(1)	1(1)	—(—)	5(9)
Internal Environment-Physical layout and Environment	—(—)	1(1)	1(1)	—(—)	—(—)	3(4)	5(6)
External Environment-Geographical factor (e.g. Location of patient)	1(1)	—(—)	—(—)	—(—)	—(—)	1(1)	2(2)
Total	49(87)	49(81)	35(55)	9(13)	34(46)	415(688)	591(970)

Each cell of the table shows the number of reports (total number of concepts) across each ethnic group for Batch 1.

DNR: Data not received, MB: Mixed Background, WO: White Other, WB: White British – denotes that the concept did not appear in the particular ethnic group's reports.

Supplementary Table 5: Batch 1 reports and their associated outcomes across the ethnic groups

Ethnic group	TH	NND	MD	Stillbirth	Total
Asian	3	1	1	1	6
Black	3	–	4	–	7
Data not received	–	–	4	–	4
Mixed Background	1	–	–	–	1
White British	30	5	8	9	52
White Other	2	–	2	2	6
Total	39	6	19	12	76

This information was only provided with Batch 1.

Therapeutic Hypothermia: TH, Early Neonatal Death: NND, Mother's death: MD.

– denotes that the outcome did not appear in the particular ethnic group's reports.

Supplementary Table 6: Performance evaluation for each test

	Test with real data		Tests with synthetic data			
	Test B		Test A		Test C	
	Avg	SD	Avg	SD	Avg	SD
Precision	0.87	0.34	1.00	0.00	1.00	0.00
Recall	0.93	0.18	0.60	0.23	0.67	0.15
F-score	0.96	0.10	0.79	0.08	0.81	0.08
Misclassification	0.19	0.35	0.40	0.23	0.33	0.15
Accuracy	0.81	0.35	0.60	0.23	0.67	0.15
Balanced Accuracy	0.90	0.18	0.80	0.11	0.83	0.08

The results of Test A and Test C can be compared because both tests were conducted on Batch 4. Test B was conducted on Batch 3 and hence cannot be directly compared to the results of Test A and Test C.

Avg: Average; SD: Standard deviation.



Supplementary Table 7: Percentage of correctly annotated concepts across all the tests A, B, and C

Concept	Test with real data	Tests with synthetic data	
	Test B	Test A	Test C
Organisation-Communication factor	82.70	45.59	62.32
Job/task-Assessment, investigation, testing, screening (e.g. holistic review)	93.39	60.97	61.26
External Environment-COVID	86.67	55.59	59.95
Organisation-Teamworking	90.53	74.40	60.58
Organisation-National guidance	94.02	78.22	73.25
Person Patient-Physical characteristics	90.42	66.81	71.98
Organisation-Documentation	86.73	79.05	79.59
Person Staff-Slip or lapse (errors that tend to happen in routine tasks that people are doing without much conscious thought)	91.27	80.00	87.78
Job/task-Care Planning	93.90	80.74	81.27
Organisation-Escalation/referral factor (including fresh eyes reviews)	92.73	61.86	57.10
Job/task-Obstetric review	95.10	48.36	52.36
Job/task-Monitoring	97.33	46.38	63.65
Technologies and Tools-issues	78.79	80.85	79.97
Job/task-Situation awareness (e.g. loss of helicopter view)	75.00	49.33	57.00
Person Staff-Decision error (errors in conscious judgements, decisions due to lack of knowledge and from misunderstanding of a situation)	92.68	37.50	35.00
Technologies and Tools (e.g. CTG, use of screening tools)-Interpretation (e.g. CTG)	100.00	51.93	27.88
Internal Environment-Acuity (e.g. capacity of the maternity unit as a whole)	83.33	12.50	11.43
Person Patient-Psychological characteristics (e.g. stress, mental health)	72.73	52.73	52.73
Job/task-Dispensing, administering	92.86	36.37	36.36
Person Staff-Training and education (e.g. attendance at antenatal classes)	77.27	96.97	90.91
Job/task-Risk assessment	98.61	58.33	72.50
Organisation-Language barrier	–	0.00	22.22
Internal Environment-Physical layout and Environment	82.61	53.33	53.33
External Environment-Geographical factor (e.g. Location of patient)	–	50.00	50.00
Organisation-Guidance factor	100.00	–	–
Person Patient-Language barrier	95.24	–	–
Organisation-Local guidance	92.11	–	–
Organisation-National and local guidance	98.00	–	–
Average	89.77	56.57	58.35
Standard Deviation	7.85	22.02	20.50

– denotes that the concept did not appear in the particular test.



Supplementary Table 8: Percentage of correctly annotated concepts in the synthetic (test) sentences for each ethnic group

Code	Asian	Black	DNR	MB	OW	WB	Avg	SD
External Environment-COVID	64.29	83.33	50.00	0.00	66.67	69.23	55.59	29.24
External Environment-Geographical factor (e.g. Location of patient)	0.00	–	–	–	–	100.00	50.00	70.71
Internal Environment-Acuity (e.g. capacity of the maternity unit as a whole)	0.00	–	0.00	–	0.00	50.00	12.50	25.00
Internal Environment-Physical layout and Environment	–	0.00	100.00	–	–	60.00	53.33	50.33
Job/task-Assessment, investigation, testing, screening (e.g. holistic review)	57.14	75.00	80.00	0.00	80.00	73.68	60.97	31.03
Job/task-Care Planning	100.00	66.67	100.00	100.00	50.00	67.74	80.74	22.02
Job/task-Dispensing, administering	–	–	0.00	–	–	72.73	36.37	51.43
Job/task-Monitoring	66.67	0.00	0.00	–	100.00	65.22	46.38	44.56
Job/task-Obstetric review	0.00	50.00	50.00	–	60.00	81.82	48.36	30.00
Job/task-Risk assessment	0.00	100.00	–	–	–	75.00	58.33	52.04
Job/task-Situation awareness (e.g. loss of helicopter view)	100.00	33.33	33.33	0.00	–	80.00	49.33	40.17
Organisation-Communication factor	40.00	50.00	50.00	0.00	60.00	73.53	45.59	25.05
Organisation-Documentation	75.00	100.00	50.00	–	100.00	70.27	79.05	21.30
Organisation-Escalation/referral factor (including fresh eyes reviews)	60.00	66.67	100.00	–	0.00	82.61	61.86	37.88
Organisation-Language barrier	0.00	–	–	0.00	0.00	–	0.00	0.00
Organisation-National guidance	83.33	62.50	66.67	100.00	80.00	76.79	78.22	13.31
Organisation-Teamworking	66.67	66.67	53.85	100.00	83.33	75.86	74.40	15.99
Person Patient-Physical characteristics	100.00	66.67	0.00	–	100.00	67.39	66.81	40.83
Person Patient-Psychological characteristics (e.g. stress, mental health)	100.00	100.00	0.00	–	0.00	63.64	52.73	50.37
Person Staff-Decision error (errors in conscious judgements, decisions due to lack of knowledge and from misunderstanding of a situation)	–	–	–	–	0.00	75.00	37.50	53.03
Person Staff-Slip or lapse (errors that tend to happen in routine tasks that people are doing without much conscious thought)	100.00	80.00	50.00	–	100.00	70.00	80.00	21.21
Person Staff-Training and education (e.g. attendance at antenatal classes)	100.00	–	100.00	–	–	90.91	96.97	5.25
Technologies and Tools (e.g. CTG, use of screening tools)-Interpretation (e.g. CTG)	–	50.00	–	–	–	53.85	51.93	2.72
Technologies and Tools-issues	100.00	100.00	66.67	100.00	50.00	68.42	80.85	21.94

Trained on real-data (Batch 1) and tested on synthetic data (Batch 4). These are the results from testing with Batch 4 (Test A). DNR: Data not received, WO: White Other, WB: White British, MB: Mixed background, SD: Standard Deviation, Avg: Average, – denotes that the concept did not appear in the particular ethnic group.



Supplementary Table 9: Test A: Training on Batch 1 and testing on Batch 4

Ethnicity	Test A			Test C		
	Correctly	Incorrectly	Correct (%)	Correctly	Incorrectly	Correct(%)
Asian	59	28	67.82	61	26	70.11
Black	54	27	66.67	54	27	66.67
Data not received	33	22	60.00	32	23	58.18
Mixed Background	9	4	69.23	9	4	69.23
Other White	32	14	69.57	34	12	73.91
White British	499	189	72.82	501	187	72.82
	Sum: 686	Sum: 284	Mean: 70.93($\pm$ 4.30)	Sum: 691	Sum: 279	Mean: 71.24 ($\pm$ 5.67)

Test C: Retraining on Batches 2 and 3 and testing on Batch 4.



Supplementary Table 10: Percentage of correctly annotated concepts in synthetic sentences for each ethnic group (after retraining with Batch 3)

Concepts	Asian	Black	DNR	MB	OW	WB	Avg	SD
External Environment-COVID	57.14	83.33	50.00	0.00	100.00	69.23	59.95	34.46
External Environment-Geographical factor (e.g. Location of patient)	0.00	–	–	–	–	100.00	50.00	70.71
Internal Environment-Acuity (e.g. capacity of the maternity unit as a whole)	0.00	0.00	0.00	–	0.00	57.14	11.43	25.56
Internal Environment-Physical layout and Environment	–	0.00	100.00	–	–	60.00	53.33	50.33
Job/task-Assessment, investigation, testing, screening (e.g. holistic review)	57.14	75.00	80.00	0.00	80.00	75.44	61.26	31.18
Job/task-Care Planning	100.00	66.67	100.00	100.00	50.00	70.97	81.27	21.68
Job/task-Dispensing, administering	–	–	0.00	–	–	72.73	36.36	51.43
Job/task-Monitoring	66.67	50.00	0.00	100.00	100.00	65.22	63.65	37.13
Job/task-Obstetric review	0.00	50.00	50.00	–	80.00	81.82	52.36	33.11
Job/task-Risk assessment	0.00	100.00	100.00	–	100.00	62.50	72.50	43.66
Job/task-Situation awareness (e.g. loss of helicopter view)	100.00	66.67	33.33	0.00	–	85.00	57.00	40.42
Organisation-Communication factor	53.33	40.00	50.00	100.00	60.00	70.59	62.32	21.09
Organisation-Documentation	100.00	75.00	50.00	–	100.00	72.97	79.59	21.06
Organisation-Escalation/referral factor (including fresh eyes reviews)	60.00	50.00	50.00	100.00	0.00	82.61	57.10	34.24
Organisation-Language barrier	0.00	–	–	0.00	0.00	88.89	22.22	44.44
Organisation-National guidance	83.33	62.50	33.33	100.00	80.00	80.36	73.25	22.90
Organisation-Teamworking	66.67	66.67	69.23	0.00	83.33	77.59	60.58	30.42
Person Patient-Physical characteristics	100.00	66.67	0.00	100.00	100.00	65.22	71.98	39.01
Person Patient-Psychological characteristics (e.g. stress, mental health)	100.00	100.00	0.00	–	0.00	63.64	52.73	50.37
Person Staff-Decision error (errors in conscious judgements, decisions due to lack of knowledge and from misunderstanding of a situation)	0.00	–	100.00	0.00	0.00	75.00	35.00	48.73
Person Staff-Slip or lapse (errors that tend to happen in routine tasks that people are doing without much conscious thought)	100.00	100.00	50.00	100.00	100.00	76.67	87.78	20.73
Person Staff-Training and education (e.g. attendance at antenatal classes)	100.00	–	100.00	–	–	72.73	90.91	15.75
Technologies and Tools (e.g. CTG, use of screening tools)-Interpretation (e.g. CTG)	–	50.00	0.00	–	0.00	61.54	27.88	32.54
Technologies and Tools-issues	100.00	100.00	66.67	100.00	50.00	63.16	79.97	22.63

Trained with Batch 1, retrained with Batches 2 and 3, and tested on Batch 4. These are the results from testing with Batch 4 (Test C).

DNR: Data not received, WO: White Other, WB: White British, MB: Mixed background, SD: Standard Deviation, Avg: Average, – denotes that the concept did not appear in the particular ethnic group.

Supplementary Table 11: Wilcoxon signed-rank test results: Comparing Tests A and C

Ethnicity	W	p
Asian	1	0.29
Black	7	0.46
Data not received	1	0.29
Mixed Background	1.5	1.00
Other White	0	0.18
White British	44	0.59

Supplementary Table 12: Sample text segments annotations by Human Annotators using the SIRch framework

Report ID	Sentence	HA-1	HA-2	HA-3
ID-1	This meant that during pregnancy the Mother did not have an anaesthetic review, referral to a dietitian, or information shared regarding the risks associated with raised BMI during pregnancy and the need for high dose folic acid.	Assessment, investigation, testing, screening (e.g. holistic review)	Escalation/referral factor (including fresh eyes reviews)	Between staff and patient (verbal)
ID-2	The abnormal CTG was not recognised and did not prompt escalation to the obstetric team, in line with local guidance.	Interpretation (e.g. CTG)	Escalation/ referral factor (including fresh eyes reviews)	Local guidance
ID-3	Obstetric reviews of the Mother whilst in labour were not carried out in line with local guidance, these would have provided an earlier opportunity for a holistic review of the Mother's risk status and care pathway.	Obstetric review	Local guidance	Assessment, investigation, testing, screening (e.g. holistic review)

HA: Human Annotator.



Supplementary Table 13: Number of reports and concept frequencies found in the reports that were utilised for training the model

Concepts	Batch 1	Batch 2	Batch 3	Total
Organisation-Teamworking	45(90)	14(81)	96(359)	155(530)
Organisation-Communication factor	59(139)	15(48)	85(290)	159(477)
Job/task-Assessment, investigation, testing, screening (e.g. holistic review)	46(93)	14(46)	90(242)	150(381)
Person Patient-Physical characteristics	39(59)	10(22)	69(240)	118(321)
Technologies and Tools (e.g. CTG, use of screening tools)-Interpretation (e.g. CTG)	12(16)	13(35)	65(146)	90(197)
Person Staff-Slip or lapse (errors that tend to happen in routine tasks that people are doing without much conscious thought)	31(44)	7(18)	61(126)	99(188)
Person Staff-Decision error (errors in conscious judgements, decisions due to lack of knowledge and from misunderstanding of a situation)	13(19)	12(27)	64(123)	89(169)
Organisation-Documentation	34(49)	7(21)	45(98)	86(168)
Organisation-Escalation/referral factor (including fresh eyes reviews)	25(38)	8(10)	64(110)	97(158)
Organisation-National and local guidance	47(89)	8(11)	37(50)	92(150)
Job/task-Obstetric review	24(31)	7(14)	58(102)	89(147)
External Environment-COVID	56(92)	2(5)	21(45)	79(142)
Organisation-National guidance	0(0)	11(18)	76(117)	87(135)
Job/task-Care Planning	28(40)	6(10)	49(82)	83(132)
Job/task-Monitoring	21(30)	8(13)	47(75)	76(118)
Technologies and Tools-issues	23(29)	9(18)	37(66)	69(113)
Job/task-Risk assessment	11(12)	4(10)	46(72)	61(94)
Job/task-Situation awareness (e.g. loss of helicopter view)	20(27)	1(2)	30(48)	51(77)
Job/task-Dispensing, administering	11(13)	4(7)	25(42)	40(62)
Internal Environment-Acuity (e.g. capacity of the maternity unit as a whole)	10(15)	3(4)	32(36)	45(55)
Organisation-Local guidance	0(0)	9(17)	27(38)	36(55)
Person Patient-Psychological characteristics (e.g. stress, mental health)	13(15)	3(5)	24(33)	40(53)
Person Staff-Training and education (e.g. attendance at antenatal classes)	11(13)	2(5)	12(22)	25(40)
Organisation-Guidance factor	0(0)	1(1)	26(38)	27(39)
Internal Environment-Physical layout and Environment	5(6)	3(4)	13(23)	21(33)
Person Patient-Language barrier	5(9)	0(0)	10(21)	15(30)
External Environment-Geographical factor (e.g. Location of patient)	2(2)	0(0)	0(0)	2(2)
Total	591(970)	181(452)	1209(2644)	1981(4066)

Batch information is provided in Supplementary Table 1. The columns show the number of reports (concept frequency).



Supplementary Table 14: Inter-rater reliability among three human annotators (HA)

RN.	HA-1	HA-2	HA-3	Agreement
1	6	6	6	5
2	8	8	8	6
3	6	6	6	5
4	10	10	11	8
5	5	5	5	5
6	5	5	5	5
7	20	20	20	16
8	10	10	7	7
9	8	8	8	7
10	6	6	6	5
11	29	29	25	17
12	7	7	6	5
13	7	7	11	6
14	6	6	5	5
15	5	5	5	5
16	5	5	4	4
17	10	10	8	6
18	13	13	15	14
19	5	5	5	5
20	5	5	6	4

The columns show the number of sentences per report coded by each HA. The last column shows the number of sentences for which there was agreement on the annotations.

RN: Report number, HA: Human Annotator.

Total number of concept agreements: 420; Total number of concepts: 524; Number of annotators: 3. IRR: 80.15%.

Supplementary Table 15: Sample of original and synthetic (test) sentences

No.	Original sentences	Synthetic sentences
1	The Mother was appropriately risk assessed at booking for midwifery led care and the pregnancy progressed without complication until the induction of her labour at 41+5 weeks which was offered in line with national guidance.	The Mother's risk was assessed when she booked for midwifery led care and her pregnancy progressed without complication until she was induced at 41+5 weeks, which was offered in line with national guidance.
2	This would not have altered the outcome for the Baby as the Baby's birthweight was within the expected range.	Even if the outcome had been different, the baby's birthweight was still within the expected range.
3	This had an impact on the Mother's experience as she did not feel able to have in depth discussions over the telephone.	This impacted the Mother's experience as she felt unable to have in depth discussions over the telephone.
4	The Trust are encouraged to improve cross boundary working practices to ensure all mothers have equal access to local information about their care pathways.	The Trust should work on improving communication and collaboration between different departments to make sure all mothers have the same access to information about their care options.
5	The Mother was given the opportunity for a discussion about her choices and IOL processes on admission to the maternity unit.	The Mother was given the opportunity to discuss her choices and IOL processes with a staff member upon admission to the maternity unit.

In the dataset only Batch 4 contained synthetic sentences, and these were only used for testing purposes and not for training the model.

1. External Environment

- a. Policy factor
- b. Societal factor
- c. Economic factor
- d. COVID
- e. Geographical factor (e.g. Location of patient)

2. Internal Environment

- a. Physical layout and Environment
- b. Acuity (e.g. capacity of the maternity unit as a whole)
- c. Availability (e.g. operating theatres)
- d. Time of day (e.g. night working or day of the week)

3. Organisation

- a. Team culture factor (e.g. patient safety culture)
- b. Incentive factor (e.g. performance evaluation)
- c. Teamworking
- d. Communication factor
 - i. Between staff
 - ii. Between staff and patient (verbal)
- e. Documentation
- f. Escalation/referral factor (including fresh eyes reviews)
- g. National and/or local guidance
- h. Language barrier

4. Jobs/Task

- a. Assessment, investigation, testing, screening (e.g. holistic review)
- b. Care planning
- c. Dispensing, administering
- d. Monitoring
- e. Risk assessment
- f. Situation awareness (e.g. loss of helicopter view)
- g. Obstetric review

5. Technologies and Tools

- a. Issues
- b. Interpretation (e.g. CTG)

6. Person

a. Patient (characteristics and performance)

i. Characteristics

- A. Physical characteristics
- B. Psychological characteristics (e.g. stress, mental health)
- C. Language competence (English)
- D. Disability (e.g. hearing problems)
- E. Training and education (e.g. attendance at antenatal classes)
- F. Record of attendance (e.g. failure to attend antenatal classes)

ii Performance

- A. Slip or lapse (errors that tend to happen in routine tasks that people are doing without much conscious thought)
- B. Decision error (errors in conscious judgements, decisions due to lack of knowledge and from misunderstanding of a situation)
- C. Intentional rule breaking (deliberately do something different from rules)

b. Staff (characteristics and performance)

i. Characteristics

- A. Physical characteristics
- B. Psychological characteristics (e.g. stress, mental health)
- C. Language competence (English)
- D. Disability (e.g. hearing problems)
- E. Training and education (e.g. attendance at ante-natal classes)
- F. Record of attendance (e.g. failure to attend antenatal classes)

ii Performance

- A. Slip or lapse (errors that tend to happen in routine tasks that people are doing without much conscious thought)
- B. Decision error (errors in conscious judgements, decisions due to lack of knowledge and from misunderstanding of a situation)
- C. Intentional rule breaking (deliberately do something different from rules)

Supplementary File 2. Synthetic dataset generation for testing purposes

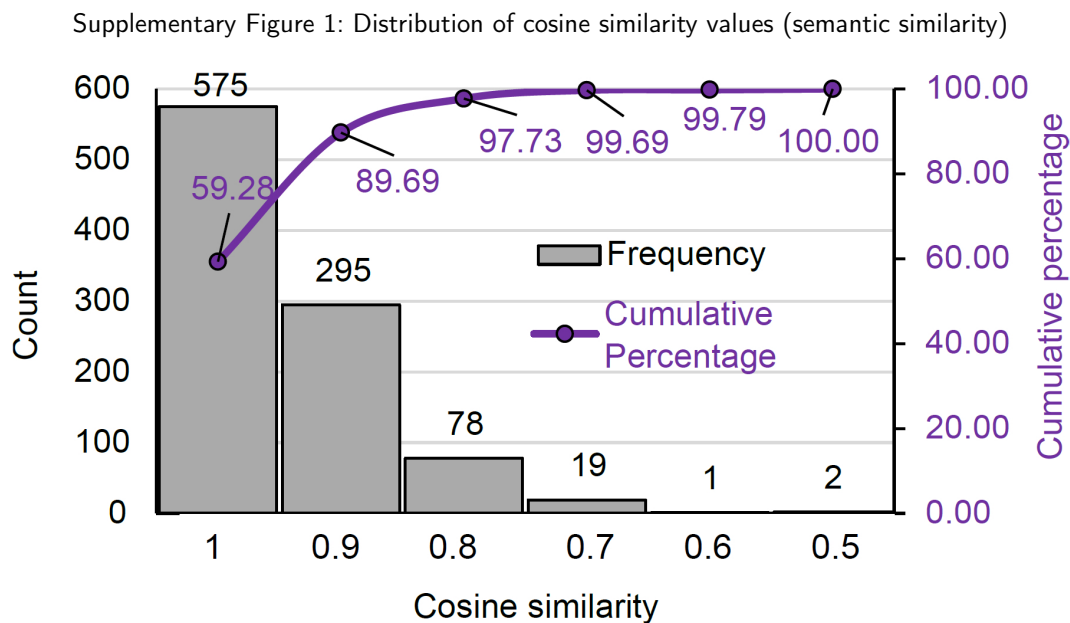
Generate synthetic sentences

To test the performance of the model on new sentences, a text generation tool was implemented that takes as input a set of original sentences from the 'Extracted text segments' and generates a new synthetic sentence per original sentence. The original and generated sentences were exported to a CSV file, with each row containing one original and its corresponding synthetic sentence. The synthetic sentences are then utilised for two purposes: 1) to further train the model by increasing its exposure to linguistic variability during training; and 2) to evaluate the performance of the trained model on new content that is semantically similar but uses different sentences. Supplementary Table 15 shows examples of original and synthetic sentences.

Sentence similarity

It is important to understand how semantically similar the original sentences are to their corresponding synthetic

sentences to evaluate the quality of the sentences generated by the synthetic sentence generation tool. To compute the similarity of each pair (i.e. original and synthetic sentence) a Word2Vec model was utilised, sentences were represented as numerical vectors based on their constituent words, capturing contextual meanings. The average vector for each sentence was then calculated by combining its word vectors. A mathematical measure, cosine similarity, was employed to compare these average sentence vectors, quantifying their semantic similarity or difference. A high cosine similarity score indicates strong semantic resemblance, while a low score suggests dissimilarity. In essence, this process allowed for the assessment of the likeness or disparity in meaning between two sentences using their word representations and cosine similarity. The similarity values of each pair were written to a CSV file. A statistical overview of the similarity values is given in Supplementary Figure 1, which shows that more than 97.5% of the sentences have a cosine similarity of more than 80% indicating good preservation of semantic content by the synthetic data generation tool. Hence, the tool effectively generated unique sentences while preserving the core meaning of the original sentence.



The higher the similarity value the closer the semantic meaning of the original and synthetic sentence (Batch 1).

