

Supplementary Materials

Predicting Chemical Accident Consequences from Substance Similarity: An On-Demand Ontology Approach

Unless stated otherwise, results refer to the leave-one-out evaluation over 1,933 accident-material cases from 1,388 accidents, with NoSignificantConsequence and Unknown excluded from the labels to be predicted, $\theta = 0.3$, cosine similarity, count aggregation, and lift ranking ($K = 3$). The R_{blind} region consists of the 207 cases (147 accidents) that record none of Injury, Fire, and Explosion. Consequence types are written as in the ontology (e.g., EnvironmentalDamage).

Supplementary Figures

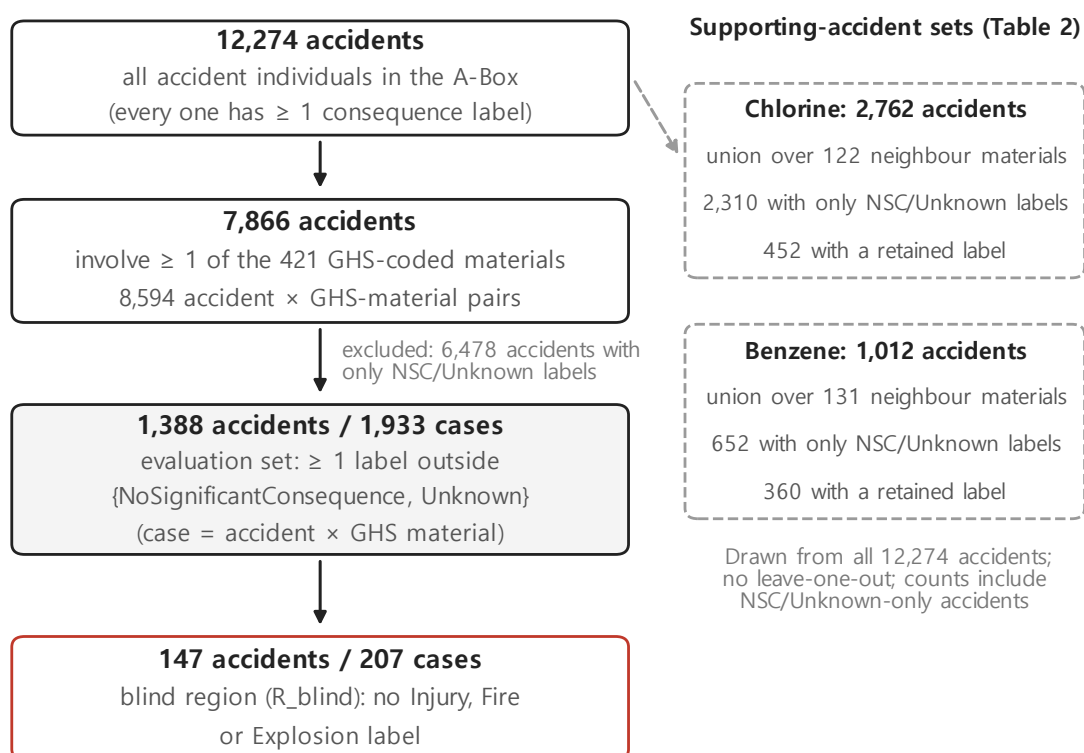


Figure S1. Data flow from the 12,274 accidents to the evaluation set and the R_{blind} region.

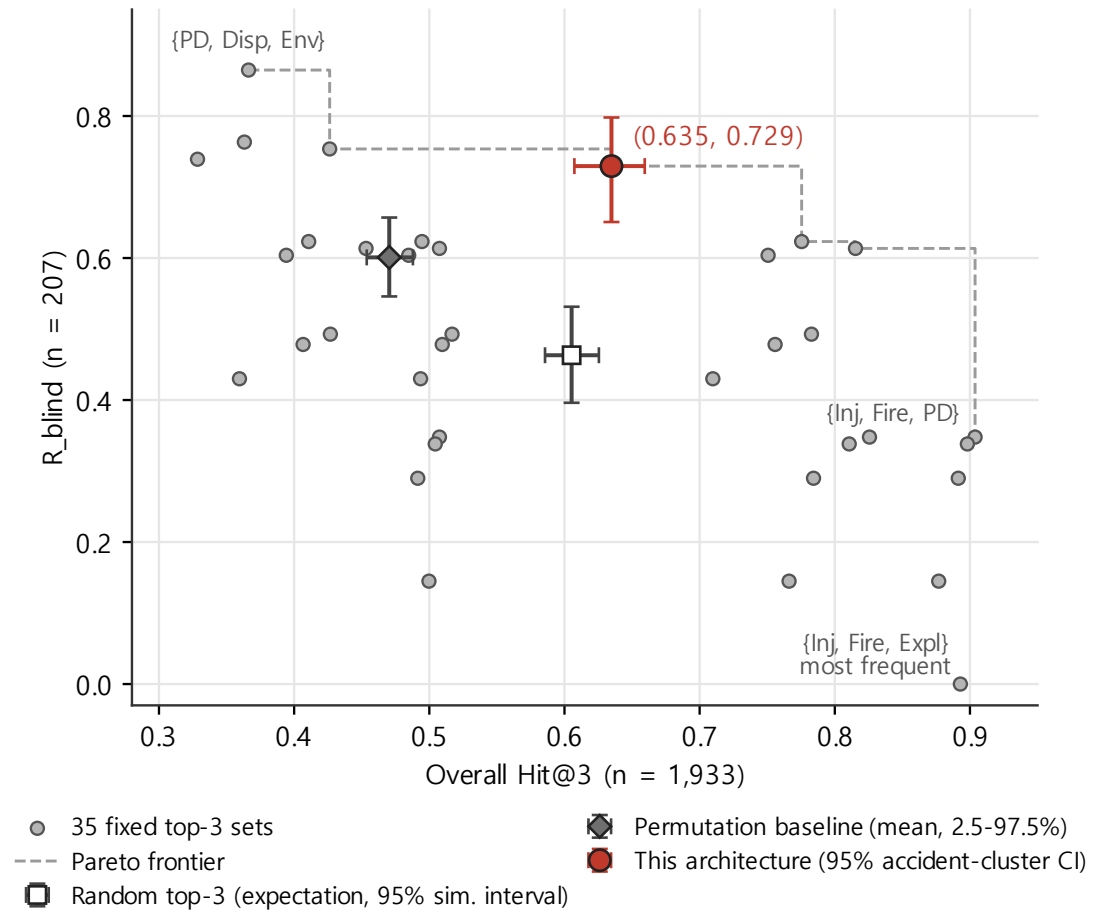


Figure S2. Overall Hit@3 against R_blind for the 35 fixed sets of three types, the random baseline, the permutation baseline, and the architecture.

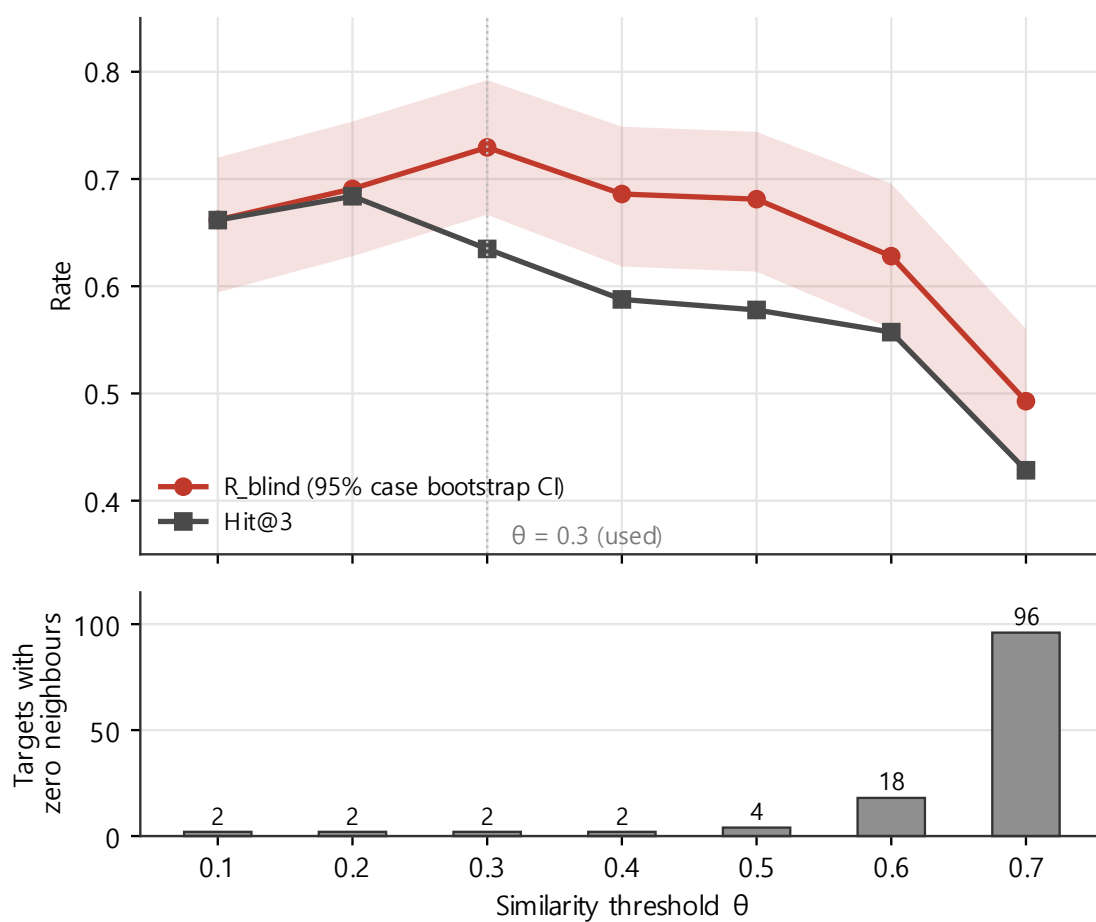


Figure S3. Sensitivity to the similarity threshold θ .

Supplementary Tables

Table S1. Data flow, accidents by source database, and consequence labels by source.

(a) Data flow from the A-Box to the evaluation set.

Step	Accidents	Cases (accident × GHS-coded material)
0. Accident individuals in the A-Box	12,274	—
1. Step 0 with ≥ 1 consequence label (hasConsequence)	12,274	—
2. Step 1 with ≥ 1 GHS-coded material (of 421)	7,866	8,594
3a. Step 2 with ≥ 1 label other than NoSignificantConsequence	1,490	2,086
3b. Step 2 with ≥ 1 label other than NoSignificantConsequence and Unknown (evaluation set)	1,388	1,933
4. R_blind region: cases of step 3b with no Injury, Fire, or Explosion label	147	207
Within step 2: accidents whose only label is NoSignificantConsequence	6,376	—
Within step 2: accidents whose only labels are NoSignificantConsequence and/or Unknown	6,478	—

Note: A case is an accident paired with one of its GHS-coded materials. The supporting data used for prediction include all accidents of the selected similar materials, also those whose only labels are NoSignificantConsequence or Unknown.

(b) Accidents by source database.

Source	Accidents	With ≥ 1 GHS-coded material	Evaluation accidents	Evaluation cases	R_blind cases (accidents)	Years (median)	Share of accidents with NoSignificantConsequence only
KHK	6,296	3,854	231	231	8 (8)	2000–2020 (2014)	0.927
KOSHA	695	570	264	308	69 (54)	2014–2022 (2017)	0.529
NTSIP	3,951	2,670	262	280	0 (0)	2013–2014 (2013)	0.905
ZEMA	157	57	18	20	1 (1)	2011–2020 (2017)	0.656
eMARS	1,175	715	613	1,094	129 (84)	1979–2022 (2003)	0.000

Note: The source is taken from the prefix of the accident identifier. The last column is the share among all accidents of the source.

(c) Number of cases per evaluation accident.

Cases per accident	Evaluation accidents
1	1,121
2	141
3	63
4	29
5	14
6	6
7	8
9	2
10	2
12	1
13	1

(d) Number of recorded consequence types per case.

Recorded types per case	All evaluation cases	R_blind cases
1	1,080	182
2	402	25
3	237	0
4	137	0
5	65	0
6	12	0

Note: Labels counted are the seven consequence types. NoSignificantConsequence and Unknown are not counted.

(e) Mapping of source fields to consequence labels (number of records with the label).

Source database	Source field	Source value (or condition)	Common label	Records
KHK	Consequence (outcome)	Injury (Injuries > 0)	Injury	443
KHK	Consequence (outcome)	Fatality (Fatalities > 0)	Fatality	27
KHK	Consequence (outcome)	NoSignificantConsequence (no casualty, no damage)	NoSignificantConsequence	5,838
NTSIP	Consequence (outcome)	Injury (Injuries > 0)	Injury	375
NTSIP	Consequence (outcome)	Fatality (Fatalities > 0)	Fatality	30
NTSIP	Consequence (outcome)	NoSignificantConsequence (no casualty, no damage)	NoSignificantConsequence	3,576
eMARS	Accident_Type (outcome)	Injury (Injuries > 0)	Injury	358
eMARS	Accident_Type (outcome)	Fatality (Fatalities > 0)	Fatality	230
eMARS	Consequence (event type)	FireAccident (event type)	Fire	682
eMARS	Consequence (event type)	ExplosionAccident (event type)	Explosion	406
eMARS	Consequence (event type)	DispersionAccident (event type)	Dispersion	194
eMARS	Accident_Type (outcome)	PropertyDamage (outcome class)	PropertyDamage	275
eMARS	Accident_Type (outcome)	EnvironmentalDamage (outcome class)	EnvironmentalDamage	86
eMARS	Accident_Type (outcome)	NoSignificantConsequence (no casualty, no damage)	NoSignificantConsequence	493
eMARS	Consequence (event type)	UnknownAccidentType (event type not given)	Unknown	310
KOSHA	Consequence (outcome)	Injury (Injuries > 0)	Injury	258
KOSHA	Consequence (outcome)	Fatality (Fatalities > 0)	Fatality	29
KOSHA	Consequence (outcome)	PropertyDamage (outcome class)	PropertyDamage	65
KOSHA	Consequence (outcome)	EnvironmentalDamage (outcome class)	EnvironmentalDamage	35
KOSHA	Consequence (outcome)	NoSignificantConsequence (no casualty, no damage)	NoSignificantConsequence	368
ZEMA	Consequence (outcome)	Injury (Injuries > 0)	Injury	52
ZEMA	Consequence (outcome)	Fatality (Fatalities > 0)	Fatality	9
ZEMA	Consequence (outcome)	NoSignificantConsequence (no casualty, no damage)	NoSignificantConsequence	103

Note: Source fields are the columns of the harmonised table of each database that was converted to the A-Box, a record may carry several labels, and the 14,242 label assignments over 12,274 accidents equal the hasConsequence counts of each source in the A-Box. In eMARS every record labelled NoSignificantConsequence (493) also carries an event type or Unknown, for example 99 carry Dispersion.

(f) Consequence labels by source database (number of accidents with the label).

Label	eMARS	KOSHA	NTSIP	KHK	ZEMA	Total
Injury	358	258	375	443	52	1,486
Fire	682	0	0	0	0	682
Explosion	406	0	0	0	0	406
PropertyDamage	275	65	0	0	0	340
Fatality	230	29	30	27	9	325
Dispersion	194	0	0	0	0	194
EnvironmentalDamage	86	35	0	0	0	121
NoSignificantConsequence	493	368	3,576	5,838	103	10,378
Unknown	310	0	0	0	0	310
Accidents (total)	1,175	695	3,951	6,296	157	12,274

Note: An accident can carry several labels, so label counts do not sum to the number of accidents. The counts cover all 12,274 accidents of the A-Box, before restriction to the evaluation set.

Table S2. Metric definitions and over-selection by type.

(a) Metric definitions.

Metric	Definition	Value (n)
Hit@3	share of cases with $T \cap P \neq \emptyset$, where T = recorded consequence types (NoSignificantConsequence and Unknown removed) and P = predicted top-3	0.635 (1,227/1,933)
Hit@3 of the most frequent prediction	Hit@3 of the fixed prediction {Injury, Fire, Explosion} (the three most frequent types)	0.893 (1,726/1,933)
R_blind region	cases with $T \cap \{\text{Injury, Fire, Explosion}\} = \emptyset$	207 cases, 147 accidents
R_blind	Hit@3 restricted to the R_blind region	0.729 (151/207)
Recovery of type c	(# cases with $c \in T$ and $c \in P$) / (# cases with $c \in T$)	main text Table 5
Precision of type c	(# cases with $c \in T$ and $c \in P$) / (# cases with $c \in P$)	main text Table 5
Prediction rate of type c	(# cases with $c \in P$) / 1,933	main text Table 5
Over-selection, neutral	prediction rate / (3/7), where 3/7 is the rate of each type if the 3 slots were filled uniformly from 7 types	Dispersion 1.99×
Over-selection, raw	prediction rate / actual rate of c among the 1,933 cases	Table S2(b)
Multilabel (example-based) Recall@3 / Precision@3 / F1@3	mean over cases of $ T \cap P / T $, $ T \cap P / P $ (0 if $P = \emptyset$), $2 T \cap P / (T + P)$	Table S10
Micro / macro P, R, F1	pooled TP/FP/FN over types (micro) or unweighted mean over the 7 types (macro, types absent in a subset are omitted from macro recall)	Table S10

(b) Over-selection by type.

Consequence type	Actual rate	Prediction rate	Over-selection, raw (prediction rate / actual rate, ×)	Over-selection, neutral (prediction rate / (3/7), ×)
------------------	-------------	-----------------	---	---

Consequence type	Actual rate	Prediction rate	Over-selection, raw (prediction rate / actual rate, ×)	Over-selection, neutral (prediction rate / (3/7), ×)
Dispersion	0.161	0.852	5.29	1.99
EnvironmentalDamage	0.067	0.558	8.36	1.30
Explosion	0.258	0.361	1.40	0.84
PropertyDamage	0.198	0.350	1.77	0.82
Injury	0.612	0.365	0.60	0.85
Fire	0.393	0.345	0.88	0.81
Fatality	0.143	0.166	1.16	0.39

Note: Rates are over the 1,933 evaluation cases. Two cases whose target has no valid code receive an empty ranking.

Table S3. Sensitivity to the threshold, the similarity measure, and the aggregation.

(a) Similarity threshold θ (cosine similarity, count aggregation, lift).

θ	Hit@3	R_blind (hits/207)	R_blind 95% CI (case-level)	Median neighbours per target	Neighbours per target: Q1–Q3	Neighbours per target: max	Targets with no neighbour (of 349)	Targets with < 5 neighbours	Targets with < 10 neighbours	Recovery: Fire	Recovery: Explosion	Recovery: Fatality	Recovery: Dispersion	Recovery: EnvironmentalDamage	Dispersion over- selection (×)
0.1	0.662	0.662 (137)	[0.594, 0.720]	295	224–340	388	2	2	2	0.362	0.233	0.000	0.981	0.822	2.28
0.2	0.684	0.691 (143)	[0.628, 0.754]	217	146–273	339	2	2	2	0.379	0.392	0.040	0.936	0.721	2.14
0.3 (used)	0.635	0.729 (151)	[0.667, 0.792]	133	75–187	275	2	2	3	0.324	0.516	0.101	0.884	0.806	1.99
0.4	0.588	0.686 (142)	[0.618, 0.749]	69	38–113	191	2	3	14	0.347	0.572	0.185	0.772	0.713	1.75
0.5	0.578	0.681 (141)	[0.614, 0.744]	29	14–51	102	4	24	55	0.447	0.608	0.120	0.794	0.597	1.76
0.6	0.557	0.628 (130)	[0.560, 0.696]	8	3–16	45	18	120	193	0.621	0.520	0.185	0.727	0.442	1.66
0.7	0.428	0.493 (102)	[0.425, 0.560]	2	0–5	28	96	258	307	0.422	0.448	0.159	0.457	0.302	1.09

(b) Similarity measure and aggregation ($\theta = 0.3$ unless stated).

Variant	Hit@3	R_blind (hits/207)	R_blind 95% CI (case-level)	Median neighbours	Recovery: Injury	Recovery: Explosion	Recovery: Fatality	Recovery: Dispersion	Recovery: EnvironmentalDamage	Dispersion over-selection (×)	Cases with a changed top-3 set (vs main analysis)
Cosine, count aggregation (used)	0.635	0.729 (151)	[0.667, 0.792]	133	0.389	0.516	0.101	0.884	0.806	1.99	—
Jaccard, $\theta = 0.3$	0.569	0.700 (145)	[0.633, 0.763]	31	0.242	0.590	0.120	0.752	0.667	1.70	1,301
Jaccard, $\theta = 0.161$ (neighbour count matched)	0.641	0.739 (153)	[0.676, 0.797]	137	0.418	0.484	0.145	0.881	0.798	1.99	519
IDF-weighted cosine (down-weights common codes)	0.672	0.667 (138)	[0.599, 0.729]	78	0.451	0.600	0.134	0.817	0.636	1.85	1,183

Variant	Hit@3	R_blind (hits/207)	R_blind 95% CI (case-level)	Median neighbours	Recovery: Injury	Recovery: Explosion	Recovery: Fatality	Recovery: Dispersion	Recovery: EnvironmentalDamage	Dispersion over-selection (×)	Cases with a changed top-3 set (vs main analysis)
IDF-weighted cosine, $\theta = 0.22$ (neighbour count matched)	0.626	0.676 (140)	[0.609, 0.739]	135	0.343	0.522	0.072	0.849	0.705	1.90	829
Aggregation weighted by the highest similarity (max-sim)	0.658	0.729 (151)	[0.667, 0.792]	133	0.422	0.566	0.072	0.878	0.752	1.96	511
Aggregation weighted by the sum of similarities (sum-sim)	0.559	0.710 (147)	[0.643, 0.773]	133	0.183	0.803	0.033	0.894	0.760	2.04	1,041

Note: CIs are percentile bootstrap intervals from 2,000 resamples of the 207 R_blind cases. Recovery is the share of cases with the type in which it is among the top three. Jaccard coefficients are discrete, so no θ gives exactly the cosine median of 133 neighbours, and $\theta = 0.161$ gives the closest median (137). Neighbour counts in (a) are over the 349 evaluated target materials, and quartiles use linear interpolation (numpy default). For the IDF-weighted cosine, $\theta = 0.22$ gives the median closest to 133 on a 0.01 grid (135), with accident-level 95% CIs of [0.600, 0.651] for Hit@3 and [0.597, 0.748] for R_blind.

Table S4. Ablation of GHS hazard-statement blocks.

Variant	Hit@3	R_blind (hits/207)	R_blind 95% CI (case- level)	Recovery: Injury	Recovery: Fire	Recovery: Explosion	Recovery: Fatality	Recovery: Dispersion	Recovery: EnvironmentalDamage (prediction rate)	R_blind, toxic	Targets with no neighbour	Cases with empty ranking	Median neighbours
All H-codes (used)	0.635	0.729 (151)	[0.667, 0.792]	0.389	0.324	0.516	0.101	0.884	0.806 (0.558)	0.846	2	2	133
Without H2xx (physical hazards)	0.416	0.623 (129)	[0.560, 0.691]	0.162	0.130	0.436	0.004	0.595	0.829 (0.622)	0.769	18	393	156
Without H3xx (health hazards)	0.752	0.589 (122)	[0.522, 0.657]	0.668	0.529	0.349	0.025	0.817	0.558 (0.399)	0.577	49	83	70
Without H4xx (environmental hazards)	0.674	0.729 (151)	[0.667, 0.787]	0.441	0.420	0.504	0.029	0.881	0.806 (0.533)	0.846	3	4	142

Note: Each block of hazard statements is removed from the code vectors before the similarity is computed. CIs are percentile bootstrap intervals from 2,000 resamples of the 207 R_blind cases. Targets with no neighbour are counted among the 349 evaluated target materials. When a block is removed, some targets lose all neighbours (18 without H2xx, 49 without H3xx, 3 without H4xx, against 2 in the main analysis), and their cases (393, 83 and 4 of the 1,933, against 2) receive an empty ranking and are counted as misses.

Table S5. Permutation, random, and fixed-set baselines.*(a) Permutation baseline.*

Metric	Observed	Permutation mean (SD)	2.5th percentile	97.5th percentile	Replicates $\geq$ observed	p (upper)
Hit@3	0.635	0.470 (0.009)	0.454	0.488	0	0.001
R_blind	0.729	0.601 (0.028)	0.546	0.657	0	0.001
Recovery, Dispersion	0.884	0.756 (0.024)	0.711	0.801	0	0.001
Recovery, EnvironmentalDamage	0.806	0.741 (0.039)	0.666	0.822	55	0.056
Recovery, Explosion	0.516	0.540 (0.021)	0.500	0.580	879	0.879
Recovery, PropertyDamage	0.470	0.450 (0.025)	0.399	0.499	227	0.228
Recovery, Injury	0.389	0.159 (0.011)	0.139	0.180	0	0.001
Recovery, Fire	0.324	0.240 (0.015)	0.212	0.271	0	0.001
Recovery, Fatality	0.101	0.055 (0.014)	0.029	0.087	0	0.001
Prediction rate, Dispersion	0.852	0.764 (0.010)	0.744	0.782	0	0.001
Prediction rate, EnvironmentalDamage	0.558	0.751 (0.010)	0.732	0.771	1000	1.000
Prediction rate, Explosion	0.361	0.547 (0.012)	0.524	0.570	1000	1.000
Prediction rate, PropertyDamage	0.350	0.467 (0.012)	0.444	0.489	1000	1.000
Prediction rate, Injury	0.365	0.159 (0.008)	0.143	0.175	0	0.001
Prediction rate, Fire	0.345	0.242 (0.010)	0.223	0.261	0	0.001
Prediction rate, Fatality	0.166	0.057 (0.005)	0.047	0.068	0	0.001
R_blind, toxic	0.846	0.696 (0.081)	0.538	0.846	43	0.044
R_blind, other	0.843	0.637 (0.061)	0.510	0.765	0	0.001
R_blind, flammable liquid	0.744	0.527 (0.070)	0.385	0.667	0	0.001
R_blind, gas	0.626	0.585 (0.039)	0.516	0.659	171	0.172

Note: 1,000 replicates. In each replicate the target material of every case is replaced by a GHS-coded material drawn uniformly from the 421 materials not involved in that accident, with the supporting data, the labels, and the R_blind region unchanged. $p = (\text{number of replicates} \geq \text{observed} + 1) / 1,001$, one-sided, so 0.001 is the smallest attainable value.

(b) Precision against the permutation baseline.

Consequence type	Precision (observed)	Permutation mean [2.5, 97.5]	Gain	p (upper)
Injury	0.652	0.612 [0.562, 0.663]	+0.039	0.056
Fire	0.369	0.390 [0.350, 0.428]	-0.021	0.856
Explosion	0.368	0.254 [0.238, 0.272]	+0.114	0.001
PropertyDamage	0.266	0.191 [0.171, 0.210]	+0.075	0.001
Fatality	0.087	0.137 [0.076, 0.203]	-0.050	0.948
Dispersion	0.167	0.159 [0.150, 0.168]	+0.008	0.042
EnvironmentalDamage	0.096	0.066 [0.059, 0.072]	+0.031	0.001

Note: Precision = recovered cases / cases with the type in the top three. Same 1,000 replicates and p-value as in (a). Gain = observed precision minus the permutation mean.

(c) Random top-3 baseline.

Subset	n	Observed	Analytic expectation	Simulated mean [2.5, 97.5]	Empirical p ($\geq$ observed)	Empirical p ($\leq$ observed)
--------	---	----------	----------------------	----------------------------	--------------------------------	--------------------------------

Subset	n	Observed	Analytic expectation	Simulated mean [2.5, 97.5]	Empirical p ($\geq$ observed)	Empirical p ($\leq$ observed)
All cases	1,933	0.635	0.605	0.605 [0.586, 0.625]	0.0017	0.9985
R_blind region	207	0.729	0.463	0.463 [0.396, 0.531]	0.0001	1.0000
R_blind region, gas	91	0.626	0.469	0.470 [0.374, 0.571]	0.0025	0.9989
R_blind region, flammable liquid	39	0.744	0.451	0.449 [0.308, 0.615]	0.0002	0.9999
R_blind region, other	51	0.843	0.451	0.451 [0.314, 0.588]	0.0001	1.0000
R_blind region, toxic	26	0.846	0.484	0.482 [0.308, 0.654]	0.0001	1.0000
All cases, gas	943	0.686	0.574	0.574 [0.545, 0.603]	0.0001	1.0000
All cases, flammable liquid	335	0.654	0.679	0.679 [0.636, 0.722]	0.8875	0.1407
All cases, other	446	0.565	0.620	0.620 [0.581, 0.659]	0.9976	0.0039
All cases, toxic	209	0.522	0.596	0.596 [0.536, 0.656]	0.9929	0.0103

Note: Three of the seven types are drawn uniformly for each case. The analytic expectation is the mean over cases of $1 - C(7 - k, 3)/C(7, 3)$, with k the number of recorded types of the case. Simulated values are from 10,000 simulations, and the empirical p-values are one-sided in each direction.

(d) The 35 fixed sets of three types.

Fixed set of three types	Hit@3 (hits/1,933)	R_blind (hits/207)	Dominated by the architecture	On the Pareto frontier of the 35 sets
PropertyDamage, Dispersion, EnvironmentalDamage	0.366 (708)	0.865 (179)	no	yes
PropertyDamage, Fatality, EnvironmentalDamage	0.363 (702)	0.763 (158)	no	no
PropertyDamage, Fatality, Dispersion	0.426 (824)	0.754 (156)	no	yes
Fatality, Dispersion, EnvironmentalDamage	0.329 (635)	0.739 (153)	no	no
Injury, PropertyDamage, EnvironmentalDamage	0.775 (1,499)	0.623 (129)	no	yes
Fire, PropertyDamage, EnvironmentalDamage	0.495 (956)	0.623 (129)	yes	no
Explosion, PropertyDamage, EnvironmentalDamage	0.411 (794)	0.623 (129)	yes	no
Injury, PropertyDamage, Dispersion	0.815 (1,576)	0.614 (127)	no	yes
Fire, PropertyDamage, Dispersion	0.508 (981)	0.614 (127)	yes	no
Explosion, PropertyDamage, Dispersion	0.453 (876)	0.614 (127)	yes	no
Injury, Dispersion, EnvironmentalDamage	0.751 (1,451)	0.604 (125)	no	no

Fixed set of three types	Hit@3 (hits/1,933)	R_blind (hits/207)	Dominated by the architecture	On the Pareto frontier of the 35 sets
Fire, Dispersion, EnvironmentalDamage	0.485 (937)	0.604 (125)	yes	no
Explosion, Dispersion, EnvironmentalDamage	0.394 (762)	0.604 (125)	yes	no
Injury, PropertyDamage, Fatality	0.783 (1,513)	0.493 (102)	no	no
Fire, PropertyDamage, Fatality	0.517 (999)	0.493 (102)	yes	no
Explosion, PropertyDamage, Fatality	0.427 (825)	0.493 (102)	yes	no
Injury, Fatality, Dispersion	0.756 (1,461)	0.478 (99)	no	no
Fire, Fatality, Dispersion	0.510 (985)	0.478 (99)	yes	no
Explosion, Fatality, Dispersion	0.407 (786)	0.478 (99)	yes	no
Injury, Fatality, EnvironmentalDamage	0.710 (1,372)	0.430 (89)	no	no
Fire, Fatality, EnvironmentalDamage	0.494 (954)	0.430 (89)	yes	no
Explosion, Fatality, EnvironmentalDamage	0.360 (695)	0.430 (89)	yes	no
Injury, Fire, PropertyDamage	0.904 (1,747)	0.348 (72)	no	yes
Injury, Explosion, PropertyDamage	0.826 (1,596)	0.348 (72)	no	no
Fire, Explosion, PropertyDamage	0.508 (981)	0.348 (72)	yes	no
Injury, Fire, Dispersion	0.898 (1,736)	0.338 (70)	no	no
Injury, Explosion, Dispersion	0.811 (1,567)	0.338 (70)	no	no
Fire, Explosion, Dispersion	0.504 (975)	0.338 (70)	yes	no
Injury, Fire, EnvironmentalDamage	0.891 (1,723)	0.290 (60)	no	no
Injury, Explosion, EnvironmentalDamage	0.784 (1,516)	0.290 (60)	no	no
Fire, Explosion, EnvironmentalDamage	0.491 (950)	0.290 (60)	yes	no
Injury, Fire, Fatality	0.877 (1,695)	0.145 (30)	no	no
Injury, Explosion, Fatality	0.766 (1,481)	0.145 (30)	no	no
Fire, Explosion, Fatality	0.500 (966)	0.145 (30)	yes	no
Injury, Fire, Explosion	0.893 (1,726)	0.000 (0)	no	no
This architecture	0.635 (1,227)	0.729 (151)	—	—

Note: No fixed set is at least as good as the architecture on both measures, and the architecture dominates 16 sets. The best probabilistic mixture of two fixed sets at Hit@3 = 0.635 reaches R_blind 0.715 (< 0.729).

Table S6. Leave-one-material-out validation.

Setting	Hit@3	R_blind (hits/207)	R_blind 95% CI (case-level)	R_blind, toxic	R_blind, other	R_blind, flammable liquid	R_blind, gas	Recovery: Injury	Recovery: Fire	Dispersion over-selection (×)
Leave-one-accident-out (main analysis)	0.635	0.729 (151)	[0.667, 0.792]	0.846	0.843	0.744	0.626	0.389	0.324	1.99
Leave-one-material-out	0.702	0.696 (144)	[0.633, 0.758]	0.846	0.843	0.667	0.582	0.563	0.243	1.99

Note: Leave-one-material-out removes every accident of the target material from the supporting data. CIs are percentile bootstrap intervals from 2,000 resamples of the 207 R_blind cases.

Table S7. Leave-one-type-out R_blind.

Type not credited	Hit@3 (hits/n)	R_blind (hits/n)	n	Permutation mean [2.5, 97.5]	p
None (main analysis)	0.635 (1,227/1,933)	0.729 (151/207)	207	0.601 [0.546, 0.657]	0.001
Injury	0.705 (848/1,203)	0.729 (151/207)	207	0.601 [0.546, 0.657]	0.001
Fire	0.622 (1,123/1,805)	0.729 (151/207)	207	0.601 [0.546, 0.657]	0.001
Explosion	0.574 (1,087/1,893)	0.729 (151/207)	207	0.601 [0.546, 0.657]	0.001
PropertyDamage	0.595 (1,118/1,879)	0.765 (117/153)	153	0.630 [0.569, 0.693]	0.001
Fatality	0.637 (1,214/1,905)	0.838 (150/179)	179	0.686 [0.626, 0.749]	0.001
Dispersion	0.574 (1,082/1,884)	0.582 (92/158)	158	0.489 [0.424, 0.557]	0.007
EnvironmentalDamage	0.618 (1,163/1,882)	0.660 (103/156)	156	0.537 [0.474, 0.603]	0.001

Note: The type is removed from the recorded labels of every case, cases left with no label are dropped, and the predictions are unchanged. n is the number of R_blind cases that keep at least one label. The permutation columns use the permutation baseline of Table S5, in which the target material of each case is replaced by a GHS-coded material drawn at random from those not involved in the accident (1,000 replicates, seed 20260909). In each replicate the same type is removed from the same cases, and the replicate's top-three lists are scored against the reduced labels. The mean and the 2.5th and 97.5th percentiles are taken over the 1,000 replicates. The one-sided p is (number of replicates reaching at least the observed R_blind + 1)/1,001, and 0.001 is the smallest attainable value.

Table S8. Timings.

(a) Stage-wise times in seconds (mean ± SD).

Configuration	n	load	similarity	injection	serialisation	OWL-RL	SHACL	retrieval	Total excl. load	Median [min, max] excl. load	Total incl. load
---------------	---	------	------------	-----------	---------------	--------	-------	-----------	------------------	------------------------------	------------------

Configuration	n	load	similarity	injection	serialisation	OWL-RL	SHACL	retrieval	Total excl. load	Median [min, max] excl. load	Total incl. load
Plan B, chlorine	10	5.517 ± 0.197	0.0001 ± 0.0000	0.301 ± 0.015	4.088 ± 0.101	2.944 ± 0.079	0.086 ± 0.005	0.0025 ± 0.0003	7.422 ± 0.184	7.414 [7.138, 7.713]	12.939 ± 0.345
Plan B, germane (not in the A-Box)	10	5.325 ± 0.066	0.0001 ± 0.0000	0.146 ± 0.003	3.874 ± 0.059	3.067 ± 0.035	0.051 ± 0.001	0.0028 ± 0.0001	7.140 ± 0.077	7.134 [7.014, 7.269]	12.465 ± 0.128
Plan C, chlorine	10	5.233 ± 0.052	0.0001 ± 0.0000	0.288 ± 0.005	3.868 ± 0.039	2.807 ± 0.026	0.082 ± 0.001	0.0025 ± 0.0001	7.048 ± 0.060	7.036 [6.971, 7.136]	12.281 ± 0.107
Plan A, precomputation	10	3.354 ± 0.109	0.0012 ± 0.0001	2.399 ± 0.080	7.876 ± 0.267	91.727 ± 2.449	—	—	102.003 ± 2.602	102.469 [98.013, 107.260]	105.357 ± 2.641
Plan A, one query	100	—	—	—	—	—	—	0.613 ± 0.023	0.613 ± 0.023	0.614 [0.565, 0.675]	0.613 ± 0.023
Plan C, cache build	10	0.015 ± 0.001	0.100 ± 0.001	—	—	—	—	—	0.100 ± 0.001	0.100 [0.098, 0.102]	0.115 ± 0.001

Note: One warm-up run discarded and 10 measured runs per configuration, run sequentially with no other Python process. For the Plan A query, each of 11 precomputed graphs was queried 11 times, and the first run and the first query of every run were discarded, leaving 100 measured queries. BLAS threads = 1 and PYTHONHASHSEED = 0 on an Intel Core i5-14600KF with 31.7 GB RAM, Windows 11 (10.0.26200), Python 3.14.6. Stage 3 of Plans B and C runs the SHACL shapes and the OWL-RL closure on the injection module with the 25 runtime axiom triples (Table S11), and all 30 measured runs of Plans B and C were accepted. For the cache build, load is reading a cached data file and similarity is computing the 421×421 cosine similarities, saving the matrix, and writing the 57,364 pairs with similarity ≥ 0 . Retrieved results were identical in every run (chlorine: 2,762 supporting accidents, top three Dispersion, Fatality, EnvironmentalDamage. Germane: 3,780 supporting accidents, top three Dispersion, Injury, Fire). A dash marks a stage that is not part of the configuration.

(b) One-off cost and per-query cost, excluding load.

Architecture	One-off cost	Per-query cost, known material	Per-query cost, unseen material
Plan A (all-pairs precomputation and global OWL-RL)	102.00 s	0.613 s	not measured (the precomputed graph has no row for the material, so the precomputation must be repeated)
Plan B (on-demand, with Stage 3)	0 s	7.422 s	7.140 s
Plan C (cached similarities, with Stage 3)	0.100 s (cache build)	7.048 s	not measured separately (same path as Plan B unseen: similarity computed on the fly + module Stage 3)

Note: Plan A's one-off cost is recovered after $102.00 / (7.422 - 0.613) = 14.98 \approx 15$ queries on known materials. Ratios of means: Plan A precomputation / Plan B chlorine = 13.74, Plan B chlorine / Plan A query = 12.12, Plan C chlorine / Plan B chlorine = 0.950, Plan B germane / Plan B chlorine = 0.962. The Plan A query has no Stage 3 because OWL-RL was applied once, globally, during precomputation.

(c) All measured runs (seconds).

Configuration	Run	Query	load	similarity	injection	serialisation	OWL-RL	SHACL	retrieval	Total excl. load	Total incl. load	Accepted by Stage 3
Plan B, chlorine	1	—	5.309229	0.000121	0.282171	3.938717	2.829173	0.085859	0.002444	7.138485	12.447714	yes
Plan B, chlorine	2	—	5.447337	0.000115	0.298537	4.038148	3.035956	0.084176	0.002376	7.459308	12.906645	yes
Plan B, chlorine	3	—	5.575812	0.000116	0.322932	4.170436	2.982911	0.080947	0.002469	7.559811	13.135623	yes
Plan B, chlorine	4	—	5.487602	0.000109	0.313221	4.102070	2.925243	0.082441	0.002442	7.425525	12.913128	yes
Plan B, chlorine	5	—	5.239093	0.000117	0.283771	4.024859	2.902200	0.096826	0.002413	7.310187	12.549280	yes
Plan B, chlorine	6	—	5.668595	0.000114	0.316155	4.236383	3.023898	0.089359	0.003075	7.668985	13.337580	yes
Plan B, chlorine	7	—	5.697966	0.000117	0.316005	4.231780	3.070187	0.092147	0.002330	7.712566	13.410532	yes
Plan B, chlorine	8	—	5.896929	0.000110	0.287650	4.113855	2.915832	0.082381	0.002961	7.402789	13.299718	yes
Plan B, chlorine	9	—	5.458448	0.000109	0.296496	4.029955	2.882490	0.083457	0.002426	7.294934	12.753382	yes
Plan B, chlorine	10	—	5.384085	0.000114	0.297372	3.991183	2.876529	0.081804	0.002518	7.249520	12.633606	yes
Plan B, germane (not in the A-Box)	1	—	5.275750	0.000097	0.146754	3.842907	3.037378	0.053075	0.002837	7.083048	12.358798	yes
Plan B, germane (not in the A-Box)	2	—	5.393497	0.000099	0.147118	3.905441	3.077465	0.051616	0.002626	7.184364	12.577861	yes
Plan B, germane (not in the A-Box)	3	—	5.408004	0.000099	0.144668	3.968445	3.101744	0.051110	0.002819	7.268884	12.676888	yes
Plan B, germane (not in the A-Box)	4	—	5.413781	0.000103	0.144447	3.857292	3.005545	0.053522	0.002741	7.063650	12.477431	yes
Plan B, germane (not in the A-Box)	5	—	5.217151	0.000084	0.142976	3.787881	3.030650	0.049243	0.002952	7.013787	12.230938	yes
Plan B, germane (not in the A-Box)	6	—	5.311674	0.000100	0.144262	3.848648	3.100234	0.050170	0.002754	7.146169	12.457843	yes
Plan B, germane (not in the A-Box)	7	—	5.342714	0.000103	0.145869	3.908557	3.096026	0.050368	0.002673	7.203596	12.546310	yes
Plan B, germane (not in the A-Box)	8	—	5.332034	0.000106	0.153176	3.956453	3.044319	0.052173	0.002744	7.208972	12.541005	yes
Plan B, germane (not in the A-Box)	9	—	5.281229	0.000088	0.143867	3.844076	3.068325	0.050133	0.002678	7.109167	12.390397	yes
Plan B, germane (not in the A-Box)	10	—	5.273263	0.000103	0.143441	3.821007	3.104686	0.049856	0.002853	7.121946	12.395209	yes
Plan C, chlorine	1	—	5.278558	0.000073	0.298975	3.893117	2.839378	0.081790	0.002396	7.115729	12.394287	yes
Plan C, chlorine	2	—	5.338791	0.000070	0.286535	3.934668	2.830419	0.082090	0.002443	7.136226	12.475017	yes
Plan C, chlorine	3	—	5.238617	0.000070	0.284080	3.841517	2.793940	0.084238	0.002533	7.006378	12.244996	yes
Plan C, chlorine	4	—	5.232168	0.000071	0.287170	3.900707	2.837865	0.083335	0.002705	7.111855	12.344023	yes
Plan C, chlorine	5	—	5.229343	0.000067	0.283636	3.883502	2.765147	0.083343	0.002321	7.018016	12.247359	yes
Plan C, chlorine	6	—	5.180446	0.000066	0.287167	3.837820	2.779266	0.081159	0.002509	6.987988	12.168433	yes
Plan C, chlorine	7	—	5.174965	0.000068	0.288410	3.801410	2.797207	0.081377	0.002503	6.970976	12.145941	yes
Plan C, chlorine	8	—	5.170136	0.000158	0.285349	3.836740	2.789115	0.083848	0.002365	6.997575	12.167711	yes
Plan C, chlorine	9	—	5.257449	0.000092	0.293791	3.884463	2.815677	0.081804	0.002378	7.078205	12.335653	yes
Plan C, chlorine	10	—	5.234419	0.000068	0.286479	3.863894	2.819568	0.081028	0.002435	7.053473	12.287892	yes
Plan A, precomputation	1	—	3.310715	0.001441	2.360912	7.857873	92.246154	0	0	102.466381	105.777095	—

Configuration	Run	Query	load	similarity	injection	serialisation	OWL-RL	SHACL	retrieval	Total excl. load	Total incl. load	Accepted by Stage 3
Plan A, one query	1	1	0	0	0	0	0	0	0.610907	0.610907	0.610907	—
Plan A, one query	1	2	0	0	0	0	0	0	0.611867	0.611867	0.611867	—
Plan A, one query	1	3	0	0	0	0	0	0	0.640201	0.640201	0.640201	—
Plan A, one query	1	4	0	0	0	0	0	0	0.615106	0.615106	0.615106	—
Plan A, one query	1	5	0	0	0	0	0	0	0.632815	0.632815	0.632815	—
Plan A, one query	1	6	0	0	0	0	0	0	0.674859	0.674859	0.674859	—
Plan A, one query	1	7	0	0	0	0	0	0	0.641286	0.641286	0.641286	—
Plan A, one query	1	8	0	0	0	0	0	0	0.614878	0.614878	0.614878	—
Plan A, one query	1	9	0	0	0	0	0	0	0.613534	0.613534	0.613534	—
Plan A, one query	1	10	0	0	0	0	0	0	0.617649	0.617649	0.617649	—
Plan A, precomputation	2	—	3.450272	0.001273	2.465261	8.002207	92.003658	0	0	102.472399	105.922671	—
Plan A, one query	2	1	0	0	0	0	0	0	0.617600	0.617600	0.617600	—
Plan A, one query	2	2	0	0	0	0	0	0	0.616394	0.616394	0.616394	—
Plan A, one query	2	3	0	0	0	0	0	0	0.599467	0.599467	0.599467	—
Plan A, one query	2	4	0	0	0	0	0	0	0.605052	0.605052	0.605052	—
Plan A, one query	2	5	0	0	0	0	0	0	0.647404	0.647404	0.647404	—
Plan A, one query	2	6	0	0	0	0	0	0	0.605569	0.605569	0.605569	—
Plan A, one query	2	7	0	0	0	0	0	0	0.651004	0.651004	0.651004	—
Plan A, one query	2	8	0	0	0	0	0	0	0.615114	0.615114	0.615114	—
Plan A, one query	2	9	0	0	0	0	0	0	0.597722	0.597722	0.597722	—
Plan A, one query	2	10	0	0	0	0	0	0	0.614618	0.614618	0.614618	—
Plan A, precomputation	3	—	3.368356	0.001122	2.368693	8.106953	92.555928	0	0	103.032696	106.401052	—
Plan A, one query	3	1	0	0	0	0	0	0	0.601986	0.601986	0.601986	—
Plan A, one query	3	2	0	0	0	0	0	0	0.607616	0.607616	0.607616	—
Plan A, one query	3	3	0	0	0	0	0	0	0.603214	0.603214	0.603214	—
Plan A, one query	3	4	0	0	0	0	0	0	0.593385	0.593385	0.593385	—
Plan A, one query	3	5	0	0	0	0	0	0	0.616880	0.616880	0.616880	—
Plan A, one query	3	6	0	0	0	0	0	0	0.603344	0.603344	0.603344	—
Plan A, one query	3	7	0	0	0	0	0	0	0.594074	0.594074	0.594074	—
Plan A, one query	3	8	0	0	0	0	0	0	0.601707	0.601707	0.601707	—
Plan A, one query	3	9	0	0	0	0	0	0	0.590360	0.590360	0.590360	—
Plan A, one query	3	10	0	0	0	0	0	0	0.630918	0.630918	0.630918	—
Plan A, precomputation	4	—	3.327640	0.001132	2.377898	7.971140	96.909777	0	0	107.259947	110.587587	—

Configuration	Run	Query	load	similarity	injection	serialisation	OWL-RL	SHACL	retrieval	Total excl. load	Total incl. load	Accepted by Stage 3
Plan A, one query	4	1	0	0	0	0	0	0	0.619778	0.619778	0.619778	—
Plan A, one query	4	2	0	0	0	0	0	0	0.613364	0.613364	0.613364	—
Plan A, one query	4	3	0	0	0	0	0	0	0.614452	0.614452	0.614452	—
Plan A, one query	4	4	0	0	0	0	0	0	0.628415	0.628415	0.628415	—
Plan A, one query	4	5	0	0	0	0	0	0	0.615483	0.615483	0.615483	—
Plan A, one query	4	6	0	0	0	0	0	0	0.607240	0.607240	0.607240	—
Plan A, one query	4	7	0	0	0	0	0	0	0.612883	0.612883	0.612883	—
Plan A, one query	4	8	0	0	0	0	0	0	0.608722	0.608722	0.608722	—
Plan A, one query	4	9	0	0	0	0	0	0	0.633204	0.633204	0.633204	—
Plan A, one query	4	10	0	0	0	0	0	0	0.615509	0.615509	0.615509	—
Plan A, precomputation	5	—	3.449572	0.001097	2.476594	8.027977	92.715527	0	0	103.221194	106.670767	—
Plan A, one query	5	1	0	0	0	0	0	0	0.616108	0.616108	0.616108	—
Plan A, one query	5	2	0	0	0	0	0	0	0.613453	0.613453	0.613453	—
Plan A, one query	5	3	0	0	0	0	0	0	0.614595	0.614595	0.614595	—
Plan A, one query	5	4	0	0	0	0	0	0	0.607288	0.607288	0.607288	—
Plan A, one query	5	5	0	0	0	0	0	0	0.601384	0.601384	0.601384	—
Plan A, one query	5	6	0	0	0	0	0	0	0.663731	0.663731	0.663731	—
Plan A, one query	5	7	0	0	0	0	0	0	0.594338	0.594338	0.594338	—
Plan A, one query	5	8	0	0	0	0	0	0	0.639665	0.639665	0.639665	—
Plan A, one query	5	9	0	0	0	0	0	0	0.603267	0.603267	0.603267	—
Plan A, one query	5	10	0	0	0	0	0	0	0.610659	0.610659	0.610659	—
Plan A, precomputation	6	—	3.517482	0.001272	2.452335	8.162366	90.527574	0	0	101.143547	104.661029	—
Plan A, one query	6	1	0	0	0	0	0	0	0.591609	0.591609	0.591609	—
Plan A, one query	6	2	0	0	0	0	0	0	0.612693	0.612693	0.612693	—
Plan A, one query	6	3	0	0	0	0	0	0	0.583258	0.583258	0.583258	—
Plan A, one query	6	4	0	0	0	0	0	0	0.593396	0.593396	0.593396	—
Plan A, one query	6	5	0	0	0	0	0	0	0.613673	0.613673	0.613673	—
Plan A, one query	6	6	0	0	0	0	0	0	0.577425	0.577425	0.577425	—
Plan A, one query	6	7	0	0	0	0	0	0	0.597626	0.597626	0.597626	—
Plan A, one query	6	8	0	0	0	0	0	0	0.628473	0.628473	0.628473	—
Plan A, one query	6	9	0	0	0	0	0	0	0.588341	0.588341	0.588341	—
Plan A, one query	6	10	0	0	0	0	0	0	0.602879	0.602879	0.602879	—
Plan A, precomputation	7	—	3.279183	0.001117	2.292726	7.672221	90.075995	0	0	100.042059	103.321242	—

Configuration	Run	Query	load	similarity	injection	serialisation	OWL-RL	SHACL	retrieval	Total excl. load	Total incl. load	Accepted by Stage 3
Plan A, one query	7	1	0	0	0	0	0	0	0.643590	0.643590	0.643590	—
Plan A, one query	7	2	0	0	0	0	0	0	0.646645	0.646645	0.646645	—
Plan A, one query	7	3	0	0	0	0	0	0	0.664206	0.664206	0.664206	—
Plan A, one query	7	4	0	0	0	0	0	0	0.586500	0.586500	0.586500	—
Plan A, one query	7	5	0	0	0	0	0	0	0.640785	0.640785	0.640785	—
Plan A, one query	7	6	0	0	0	0	0	0	0.630453	0.630453	0.630453	—
Plan A, one query	7	7	0	0	0	0	0	0	0.616168	0.616168	0.616168	—
Plan A, one query	7	8	0	0	0	0	0	0	0.615861	0.615861	0.615861	—
Plan A, one query	7	9	0	0	0	0	0	0	0.646139	0.646139	0.646139	—
Plan A, one query	7	10	0	0	0	0	0	0	0.614742	0.614742	0.614742	—
Plan A, precomputation	8	—	3.313746	0.001073	2.429375	7.774840	87.808188	0	0	98.013477	101.327223	—
Plan A, one query	8	1	0	0	0	0	0	0	0.565897	0.565897	0.565897	—
Plan A, one query	8	2	0	0	0	0	0	0	0.566465	0.566465	0.566465	—
Plan A, one query	8	3	0	0	0	0	0	0	0.573646	0.573646	0.573646	—
Plan A, one query	8	4	0	0	0	0	0	0	0.564658	0.564658	0.564658	—
Plan A, one query	8	5	0	0	0	0	0	0	0.565274	0.565274	0.565274	—
Plan A, one query	8	6	0	0	0	0	0	0	0.574757	0.574757	0.574757	—
Plan A, one query	8	7	0	0	0	0	0	0	0.577462	0.577462	0.577462	—
Plan A, one query	8	8	0	0	0	0	0	0	0.566095	0.566095	0.566095	—
Plan A, one query	8	9	0	0	0	0	0	0	0.575923	0.575923	0.575923	—
Plan A, one query	8	10	0	0	0	0	0	0	0.565097	0.565097	0.565097	—
Plan A, precomputation	9	—	3.130092	0.001153	2.261467	7.243372	89.624794	0	0	99.130786	102.260878	—
Plan A, one query	9	1	0	0	0	0	0	0	0.626711	0.626711	0.626711	—
Plan A, one query	9	2	0	0	0	0	0	0	0.610021	0.610021	0.610021	—
Plan A, one query	9	3	0	0	0	0	0	0	0.634681	0.634681	0.634681	—
Plan A, one query	9	4	0	0	0	0	0	0	0.619410	0.619410	0.619410	—
Plan A, one query	9	5	0	0	0	0	0	0	0.609656	0.609656	0.609656	—
Plan A, one query	9	6	0	0	0	0	0	0	0.626619	0.626619	0.626619	—
Plan A, one query	9	7	0	0	0	0	0	0	0.672305	0.672305	0.672305	—
Plan A, one query	9	8	0	0	0	0	0	0	0.639025	0.639025	0.639025	—
Plan A, one query	9	9	0	0	0	0	0	0	0.628780	0.628780	0.628780	—
Plan A, one query	9	10	0	0	0	0	0	0	0.636245	0.636245	0.636245	—
Plan A, precomputation	10	—	3.392521	0.001236	2.501760	7.937373	92.806084	0	0	103.246454	106.638974	—

Configuration	Run	Query	load	similarity	injection	serialisation	OWL-RL	SHACL	retrieval	Total excl. load	Total incl. load	Accepted by Stage 3
Plan A, one query	10	1	0	0	0	0	0	0	0.621262	0.621262	0.621262	—
Plan A, one query	10	2	0	0	0	0	0	0	0.629741	0.629741	0.629741	—
Plan A, one query	10	3	0	0	0	0	0	0	0.617664	0.617664	0.617664	—
Plan A, one query	10	4	0	0	0	0	0	0	0.605467	0.605467	0.605467	—
Plan A, one query	10	5	0	0	0	0	0	0	0.603791	0.603791	0.603791	—
Plan A, one query	10	6	0	0	0	0	0	0	0.619739	0.619739	0.619739	—
Plan A, one query	10	7	0	0	0	0	0	0	0.607808	0.607808	0.607808	—
Plan A, one query	10	8	0	0	0	0	0	0	0.614101	0.614101	0.614101	—
Plan A, one query	10	9	0	0	0	0	0	0	0.613917	0.613917	0.613917	—
Plan A, one query	10	10	0	0	0	0	0	0	0.613648	0.613648	0.613648	—
Plan C, cache build	1	—	0.016699	0.098400	—	—	—	—	—	0.098400	0.115099	—
Plan C, cache build	2	—	0.014827	0.098753	—	—	—	—	—	0.098753	0.113580	—
Plan C, cache build	3	—	0.014964	0.099899	—	—	—	—	—	0.099899	0.114862	—
Plan C, cache build	4	—	0.015599	0.099384	—	—	—	—	—	0.099384	0.114983	—
Plan C, cache build	5	—	0.016232	0.100226	—	—	—	—	—	0.100226	0.116458	—
Plan C, cache build	6	—	0.014928	0.099138	—	—	—	—	—	0.099138	0.114066	—
Plan C, cache build	7	—	0.015040	0.101061	—	—	—	—	—	0.101061	0.116101	—
Plan C, cache build	8	—	0.015330	0.101673	—	—	—	—	—	0.101673	0.117002	—
Plan C, cache build	9	—	0.015958	0.100249	—	—	—	—	—	0.100249	0.116207	—
Plan C, cache build	10	—	0.014676	0.099051	—	—	—	—	—	0.099051	0.113727	—

Note: A zero marks a stage that the configuration does not execute, and a dash an empty entry. Measured rows: 10 for each configuration and 100 for the Plan A query (150 in total).

Table S9. Definitions of the five ranking functions.

(a) Scores.

Function	Score $s(c)$	Source
M1 lift (used)	$s(c) = \frac{\text{obs}_t(c)}{\text{base}(c)}$	[41]
M2 log lift	$s(c) = \ln \frac{\text{obs}_t(c)}{\text{base}(c)}$ (monotone in M1, identical ranking)	[41]
M3 PMI normalised by support	$s(c) = \frac{\ln(\text{obs}_t(c)/\text{base}(c))}{-\ln(n_t(c)/N_{\text{corpus}})}$	[54], with the joint term $n_t(c)/N_{\text{corpus}}$ in place of the joint probability of standard NPMI
M4 geometric mean	$s(c) = \frac{\text{obs}_t(c)}{\sqrt{\text{base}(c)}}$	defined for this study
M5 rank displacement	$s(c) = \rho(c) - r_t(c)$. $\rho(c)$ = rank of c by base (Injury 1, Fire 2, Explosion 3, PropertyDamage 4, Fatality 5, Dispersion 6, EnvironmentalDamage 7). $r_t(c)$ = rank among candidates by $(-\text{obs}, \text{name})$	defined for this study

Note: Target t , held-out accident e . $A_t^{(-e)}$ = union of the accidents of neighbour materials ($\text{sim} \geq \theta = 0.3$, t excluded) minus e . $N_t = |A_t^{(-e)}|$ (accidents with only excluded labels included). $n_t(c) = |\{a \in A_t^{(-e)} : c \in L(a)\}|$. $\text{obs}_t(c) = n_t(c)/N_t$. $\text{base}(c) = |\{a : c \in L(a)\}|/N_{\text{corpus}}$, $N_{\text{corpus}} = 12,274$. Candidates: non-excluded types with $n_t(c) > 0$. If $N_t = 0$ the list is empty. Top- K ($K = 3$): sort by $(-s(c), -\text{obs}_t(c), c)$, with no smoothing constants. Fewer than 3 candidates give a shorter list (no padding). In M3 the joint term is $n_t(c)/N_{\text{corpus}}$, which differs from the standard NPMI $p(t, c)$. PMI normalised by support is therefore not the standard normalised PMI.

(b) Case-level bootstrap CIs for the five ranking functions (main text Table 6).

Ranking function	R_blind 95% CI (all 1,933 cases resampled)	R_blind 95% CI (207 R_blind cases resampled)	Dispersion over-selection 95% CI	Largest over-selection 95% CI
lift (M1, used here)	[0.668, 0.788]	[0.667, 0.792]	[1.95, 2.02]	[1.95, 2.02]
log lift (M2)	[0.668, 0.788]	[0.667, 0.792]	[1.95, 2.02]	[1.95, 2.02]
PMI normalised by support (M3)	[0.596, 0.725]	[0.594, 0.725]	[1.87, 1.95]	[1.87, 1.95]
geometric mean (M4)	[0.289, 0.418]	[0.285, 0.420]	[1.22, 1.32]	[2.30, 2.32]
rank displacement (M5)	[0.308, 0.438]	[0.304, 0.440]	[1.66, 1.75]	[2.07, 2.13]

Note: Percentile bootstrap intervals from 2,000 resamples of cases. Point values and accident-level CIs are in main text Table 6.

Table S10. Multilabel metrics at $K = 3$.

(a) Multilabel metrics.

Subset	n	Hit@3	Multilabel (example-based) Recall@3 / Precision@3 / F1@3	Micro P / R / F1	Macro P / R / F1	TP / FP / FN	Mean labels per case
All cases	1,933	0.635	0.450 / 0.267 / 0.310	0.268 / 0.438 / 0.332	0.286 / 0.499 / 0.307	1,550 / 4,243 / 1,990	1.831

Subset	n	Hit@3	Multilabel (example-based) Recall@3 / Precision@3 / F1@3	Micro P / R / F1	Macro P / R / F1	TP / FP / FN	Mean labels per case
R_blind region	207	0.729	0.691 / 0.258 / 0.370	0.258 / 0.690 / 0.375	0.199 / 0.604 / 0.243	160 / 461 / 72	1.121

Note: Macro recall averages over the types present in the subset (7 overall, 4 in the R_blind region).

(b) Hit@3 by number of recorded consequence types.

Subset	Recorded types per case	n	Hits	Hit@3
All cases	1	1,080	493	0.456
All cases	2	402	321	0.799
All cases	3	237	210	0.886
All cases	4	137	127	0.927
All cases	5	65	64	0.985
All cases	6	12	12	1.000
R_blind region	1	182	126	0.692
R_blind region	2	25	25	1.000

Table S11. Runtime axioms, SHACL shapes, validation module, and fault-injection results for Stage 3.

(a) Runtime axioms for the injected vocabulary (25 triples, Turtle).

```

@prefix : <http://www.semanticweb.org/user/ontologies/2025/8/untitled-ontology-50#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

:similarTo a owl:IrreflexiveProperty,
  owl:ObjectProperty ;
  owl:propertyChainAxiom ( :isSubjectMaterialOf :hasSimilarMaterial ) .

:similarityScore a owl:DatatypeProperty,
  owl:FunctionalProperty ;
  rdfs:domain :MaterialSimilarity ;
  rdfs:range xsd:decimal .

:hasSimilarMaterial a owl:FunctionalProperty,
  owl:ObjectProperty ;
  rdfs:domain :MaterialSimilarity ;
  rdfs:range :Material .

:hasSubjectMaterial a owl:FunctionalProperty,
  owl:ObjectProperty ;
  rdfs:domain :MaterialSimilarity ;
  rdfs:range :Material .

:isSubjectMaterialOf a owl:ObjectProperty ;
  owl:inverseOf :hasSubjectMaterial .

:Material owl:disjointWith :Accident .

:MaterialSimilarity a owl:Class ;
  owl:disjointWith :Accident,
    :Material .

```

Note: The axioms are added in memory at query time and are never written to the T-Box. The property chain makes `:similarTo` link a subject material to its similar material, and its irreflexivity forbids a material being declared similar to itself.

(b) SHACL shapes (Turtle).

```
@prefix sh: <http://www.w3.org/ns/shacl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix : <http://www.semanticweb.org/user/ontologies/2025/8/untitled-ontology-50#> .
@prefix ds: <urn:ondemand-ontology:shapes#> .

ds:MaterialSimilarityShape a sh:NodeShape ;
  sh:targetClass :MaterialSimilarity ;
  sh:targetSubjectsOf :hasSubjectMaterial , :hasSimilarMaterial , :similarityScore ;
  sh:class :MaterialSimilarity ;
  sh:not [ sh:class :Material ] ;
  sh:not [ sh:class :Accident ] ;
  sh:property [ sh:path :hasSubjectMaterial ; sh:minCount 1 ; sh:maxCount 1 ;
    sh:class :Material ; sh:nodeKind sh:IRI ;
    sh:disjoint :hasSimilarMaterial ] ;
  sh:property [ sh:path :hasSimilarMaterial ; sh:minCount 1 ; sh:maxCount 1 ;
    sh:class :Material ; sh:nodeKind sh:IRI ] ;
  sh:property [ sh:path :similarityScore ; sh:minCount 1 ; sh:maxCount 1 ;
    sh:or ( [ sh:datatype xsd:decimal ] [ sh:datatype xsd:float ] [ sh:datatype
xsd:double ] ) ;
    sh:minInclusive 0 ; sh:maxInclusive 1 ] .

ds:SameAsDifferentFromShape a sh:NodeShape ;
  sh:targetSubjectsOf owl:sameAs ;
  sh:sparql [
    a sh:SPARQLConstraint ;
    sh:message "owl:sameAs and owl:differentFrom asserted on the same pair" ;
    sh:prefixes ds:prefixes ;
    sh:select """SELECT $this ?value WHERE {
      $this owl:sameAs ?value . FILTER($this != ?value)
      { $this owl:differentFrom ?value } UNION { ?value owl:differentFrom $this } }""" ] .

ds:prefixes sh:declare [ sh:prefix "owl" ; sh:namespace
"http://www.w3.org/2002/07/owl#"^^xsd:anyURI ] .
```

Note: The shapes are validated with pySHACL on the asserted module, before the OWL-RL closure and without inference.

(c) Definition of the validation module.

For an injection Δ , the validation module $M(\Delta)$ of Stage 3 is the union of four sets of triples: (1) the T-Box; (2) the 25 runtime axiom triples of (a); (3) every asserted triple whose subject is an injected node, that is, a resource that is the subject of `rdf:type :MaterialSimilarity`, `:hasSubjectMaterial`, `:hasSimilarMaterial`, or `:similarityScore`; and (4) every asserted triple whose subject is an IRI occurring as the object of such a triple, namely the subject material, the similar material, and the class IRIs. The module thus extends exactly one hop from the injected nodes, and literal objects are kept as values but not expanded. All `owl:sameAs` and `owl:differentFrom` assertions of the whole graph are also included. Accident individuals and all other A-Box triples are excluded unless they lie one hop from an injected node.

(d) Fault-injection results.

Case	Injected content	Injected nodes	Module triples	OWL-RL (module)	SHACL (module)	SHACL constraint components	Decision (module)	Whole-graph triples	Decision (whole graph)	Same decision	Same detectors	Same new messages
Well-formed (Chlorine)	Correct injection, 122 neighbours	122	4,999	passed	passed	—	accepted	271,765	accepted	yes	yes	yes
Well-formed (Benzene)	Correct injection, 131 neighbours	131	3,213	passed	passed	—	accepted	271,810	accepted	yes	yes	yes
Well-formed (Germane, not in the A-Box)	Correct injection for the query codes H220, H280, H302, and H330, 61 neighbours	61	4,907	passed	passed	—	accepted	271,463	accepted	yes	yes	yes
F1	Two similar materials asserted to be both owl:sameAs and owl:differentFrom each other	122	5,001	detected	detected	SPARQL	rejected	271,767	rejected	yes	yes	yes
F2	Similarity score of one node set to 1.5	122	4,999	passed	detected	MaxInclusive	rejected	271,765	rejected	yes	yes	yes
F3	Extra node whose subject and similar material are both the target, score 1.0	123	5,003	detected	detected	Disjoint	rejected	271,769	rejected	yes	yes	yes
F4	Second similarity score (0.1234) added to one node	122	5,000	passed	detected	MaxCount	rejected	271,766	rejected	yes	yes	yes
F5	Second similar material (hymexazol, not a neighbour) added to one node	122	5,011	passed	detected	MaxCount	rejected	271,766	rejected	yes	yes	yes
F6	One node additionally typed as a material	122	5,000	detected	detected	Not	rejected	271,766	rejected	yes	yes	yes
F7	Similar material of one node replaced by an accident individual	122	4,933	detected	detected	Class	rejected	271,765	rejected	yes	yes	yes
F8	Similarity score of one node given as the non-numeric decimal "high"	122	4,999	detected	detected	MinInclusive, Or	rejected	271,765	rejected	yes	yes	yes
F9	Subject material of one node removed	122	4,998	passed	detected	MinCount	rejected	271,764	rejected	yes	yes	yes

Case	Injected content	Injected nodes	Module triples	OWL-RL (module)	SHACL (module)	SHACL constraint components	Decision (module)	Whole-graph triples	Decision (whole graph)	Same decision	Same detectors	Same new messages
F10	Similarity score of one node given as the string "0.9"	122	4,999	detected	detected	MaxInclusive, MinInclusive, Or	rejected	271,765	rejected	yes	yes	yes

Note: The ten faults were introduced into the well-formed chlorine injection (122 nodes). Module = the scope defined in (c). Whole graph = the same two checks applied to the full snapshot. An injection is rejected if either check reports a message that is absent from the baseline run on the graph without injection. The last three columns compare the module with the whole graph for the decision, for which check detected the fault, and for the new messages themselves.

Table S12. Fold-wise base rates.

Quantity	Static base rates (main analysis)	Fold-wise base rates
Hit@3 (hits)	0.635 (1,227)	0.636 (1,230)
R_blind	0.729 (151/207)	0.729 (151/207)
R_blind 95% CI (case-level)	[0.667, 0.792]	[0.667, 0.792]
R_blind 95% CI (accident-level)	[0.651, 0.798]	[0.651, 0.798]
Recovery: Injury	0.389	0.389
Recovery: Fire	0.324	0.333
Recovery: Explosion	0.516	0.516
Recovery: PropertyDamage	0.470	0.470
Recovery: Fatality	0.101	0.105
Recovery: Dispersion	0.884	0.887
Recovery: EnvironmentalDamage	0.806	0.806
Dispersion over-selection (×)	1.99	1.99
R_blind, toxic / other / flammable liquid / gas	0.846 / 0.843 / 0.744 / 0.626	0.846 / 0.843 / 0.744 / 0.626
Most frequent prediction	{Injury, Fire, Explosion}	{Injury, Fire, Explosion} in all 1,933 folds
R_blind region	207 cases	identical (207 cases)
Cases whose top-3 set changed	—	10
Maximum absolute change of any base rate	—	8.07×10^{-5}

Note: Fold-wise: the base rates of lift and the most frequent prediction are recomputed in each leave-one-out fold without the withheld accident. Static: both are computed once over all 12,274 accidents.