

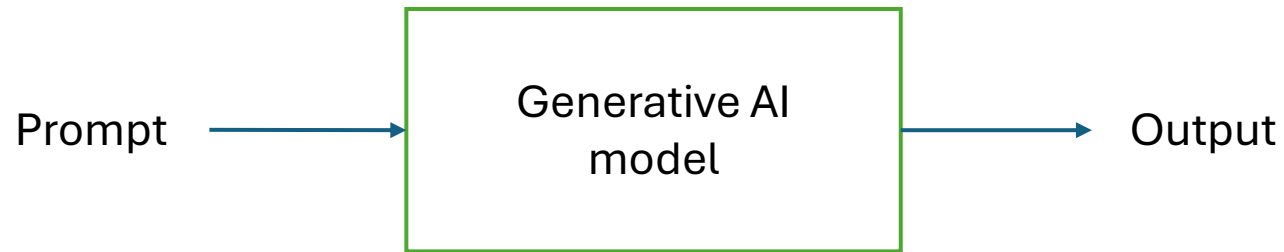
Generative AI for Research Data Processing: Lessons Learnt From Three Use Cases

Modhurita Mitra

Research Engineering Team
Research and Data Management Services
Utrecht University


Work done jointly with Martine de Vos, Nicola Cortinovis, and Dawa Ometto

Generative AI : General-purpose technology



Prompts in natural language → Low barrier to entry for domain researchers

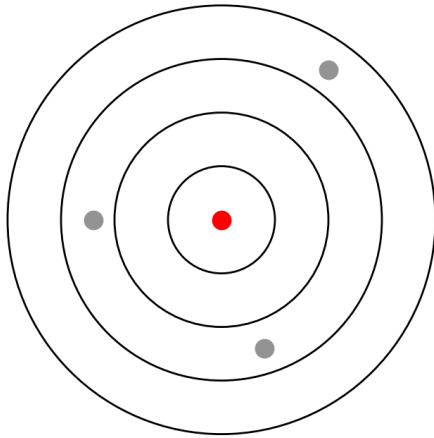


Andrej Karpathy  @karpathy · Jan 24, 2023

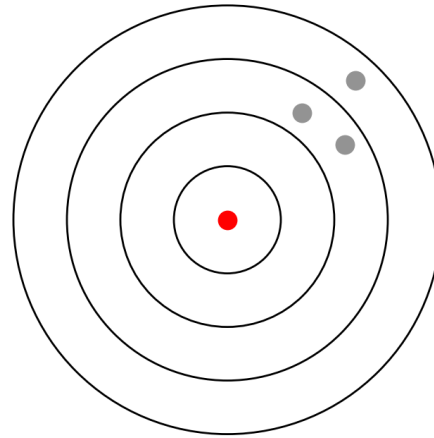
The hottest new programming language is English

Goals of the study

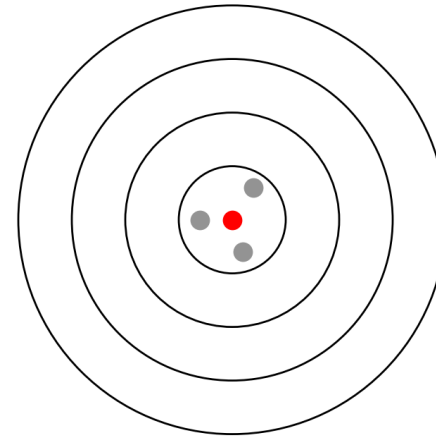
- Appropriateness
- Accuracy and consistency



Neither Accurate nor Consistent



Not Accurate but Consistent



Both Accurate and Consistent

Use cases

Project	Scientific domain	Task	Type of task	Amount of data
Seedlists	Botanical Gardens	Extract plant species names from historical seedlists (catalogues of seeds) published by botanical gardens	Information extraction	~30,000 seedlists
Health Technology Assessment (HTA) documents	Pharmaceutical Sciences	Extract certain data points from documents published by HTA organizations in the EU	Natural language understanding	A few thousand HTA documents
Kickstarter	Economic Geography	Assign industry codes to projects on the crowdfunding website Kickstarter	Text classification	~300,000 Kickstarter projects

Use case: Seedlists

- 7 -

Salvia austriaca L.	Scilla amoena Horn.	Serratula coronata L.	Silene vesperitia Retz.
— ciliatissima L.	— italica L.	— heterophylla Desf.	— verna Pers.
— dormans Lam.	— verna Huds.	— quinquifolia M. Bieb.	— Zaveland Herb.
— Farskoldi Vahl.	Scirpus atrovirens Willd.	— radiata M. B.	— pubescens Lam.
— lanata Jacq.	Scirpus amurensis Schrad.	— Scallii graveolens Ledeb.	— Siamensis austricum Jacq.
— hispanica L.	Scorpioides dactylota L.	— gumifera Sm.	— brachycarpum F. et M.
— linaria C. A. M.	Scorzonera austriaca Willd.	— Pilsnii Bess.	— Camptocarpum F. et M.
— micrantha Vahl.	— crinitus M. Bieb.	— humilis L.	— luteum Retz.
— multifida Sibth.	— villosa Scop.	— sphaerocephala Ard.	— luteum Retz.
— pteris R. Br.	Scrophularia Bolanderi Horn.	— Steyeri Benth.	— luteum Retz.
— pseudocrocata Jacq.	— canalis L.	— tenuifolia Schrad.	— luteum Retz.
— rhomboides R. et P.	— Scutellaria alpina L.	— aurita Willd.	— luteum Retz.
— rubra Ehrh.	— lutea L.	— crissa Lam.	— luteum Retz.
— sylvatica Burch.	— peregrina L.	— glauca Cav.	— luteum Retz.
— verbasifolia M. Bieb.	— scordifolia Fisch.	— mollissima Willd.	— luteum Retz.
— Verbenae L.	— Solanum angustifolium Schrad.	— thymifolia Lam.	— luteum Retz.
— virgata Ait.	— Solanum angustifolium Schrad.	— Sellowiana Klotzsch.	— luteum Retz.
— viridis L.	— Solanum angustifolium Schrad.	— spinosa L.	— luteum Retz.
— sanguinea media L.	— Solanum angustifolium Schrad.	— spiraeifolia Lk.	— luteum Retz.
— tenuifolia Fisch.	— Solanum angustifolium Schrad.	— stratum Sm.	— luteum Retz.
— Sanicula maritima L.	— Solanum angustifolium Schrad.	— ramosa L.	— luteum Retz.
— Santolina incana Lam.	— Solanum angustifolium Schrad.	— scordifolia L.	— luteum Retz.
— leucantha Bertol.	— Solanum angustifolium Schrad.	— taurica M. Bieb.	— luteum Retz.
— pinnata Vitr.	— Solanum angustifolium Schrad.	— Steyeri Benth.	— luteum Retz.
— aquatica Willd.	— Solanum angustifolium Schrad.	— Boissiana Bertol.	— luteum Retz.
— viridis Willd.	— Solanum angustifolium Schrad.	— orientalis Lam.	— luteum Retz.
— Saponaria cerastoides C. A. Meyer.	— Solanum angustifolium Schrad.	— Sieversii pyramidalis Rehb.	— luteum Retz.
— glauca M. Bieb.	— Solanum angustifolium Schrad.	— Silesiacensis Jacq.	— luteum Retz.
— ocyroides L.	— Solanum angustifolium Schrad.	— annulata Thor.	— luteum Retz.
— portigena Lam.	— Solanum angustifolium Schrad.	— apica Turca.	— luteum Retz.
— viscosa F. et M.	— Solanum angustifolium Schrad.	— Ancon Jacq.	— luteum Retz.
— Sarracenia coccinea R. et P.	— Solanum angustifolium Schrad.	— Belen L.	— luteum Retz.
— Sarracenia coccinea R. et P.	— Solanum angustifolium Schrad.	— brachyptera Rob. Cast.	— luteum Retz.
— Satureja cuneifolia Ten.	— Solanum angustifolium Schrad.	— pallidum M. B.	— luteum Retz.
— intermedia C. A. M.	— Solanum angustifolium Schrad.	— populifolium Pall.	— luteum Retz.
— Thymus L.	— Solanum angustifolium Schrad.	— pulchrum Michx.	— luteum Retz.
— Saussurea alpina Cand.	— Solanum angustifolium Schrad.	— scordifolia M. Bieb.	— luteum Retz.
— elongata Cand.	— Solanum angustifolium Schrad.	— Sieboldii Mark.	— luteum Retz.
— pulchella Fisch.	— Solanum angustifolium Schrad.	— spiricum M. B.	— luteum Retz.
— salicifolia Cand.	— Solanum angustifolium Schrad.	— fasciculata Lam.	— luteum Retz.
— Saxifraga aizoides L.	— Solanum angustifolium Schrad.	— spuria L.	— luteum Retz.
— austriaca Lam.	— Solanum angustifolium Schrad.	— githium Ait.	— luteum Retz.
— bulbifera L.	— Solanum angustifolium Schrad.	— Smithii Sims.	— luteum Retz.
— cernua L.	— Solanum angustifolium Schrad.	— spathulatum Horn.	— luteum Retz.
— controversa Sternb.	— Solanum angustifolium Schrad.	— tabulaforme Hae.	— luteum Retz.
— cuneifolia L.	— Solanum angustifolium Schrad.	— tortuosum Ait.	— luteum Retz.
— exarata Vill.	— Solanum angustifolium Schrad.	— Smithii Sims.	— luteum Retz.
— graminifolia L.	— Solanum angustifolium Schrad.	— spathulatum Horn.	— luteum Retz.
— irrigua M. B.	— Solanum angustifolium Schrad.	— tabulaforme Hae.	— luteum Retz.
— leptophylla Pers.	— Solanum angustifolium Schrad.	— tortuosum Ait.	— luteum Retz.
— muscoides Willd.	— Solanum angustifolium Schrad.	— Smithii Sims.	— luteum Retz.
— mutata L.	— Solanum angustifolium Schrad.	— spathulatum Horn.	— luteum Retz.
— nitida L.	— Solanum angustifolium Schrad.	— tabulaforme Hae.	— luteum Retz.
— pedatifida Ehrh.	— Solanum angustifolium Schrad.	— tortuosum Ait.	— luteum Retz.
— peruviana Lam.	— Solanum angustifolium Schrad.	— Smithii Sims.	— luteum Retz.
— planifolia L.	— Solanum angustifolium Schrad.	— spathulatum Horn.	— luteum Retz.
— alberta L.	— Solanum angustifolium Schrad.	— tabulaforme Hae.	— luteum Retz.
— stellata L.	— Solanum angustifolium Schrad.	— tortuosum Ait.	— luteum Retz.
— tenella Willd.	— Solanum angustifolium Schrad.	— Smithii Sims.	— luteum Retz.
— trifurcata Schrad.	— Solanum angustifolium Schrad.	— spathulatum Horn.	— luteum Retz.
— Scabiosa grammata L.	— Solanum angustifolium Schrad.	— tabulaforme Hae.	— luteum Retz.
— Hylaeus Lam.	— Solanum angustifolium Schrad.	— tortuosum Ait.	— luteum Retz.
— holosericea Bertol.	— Solanum angustifolium Schrad.	— Smithii Sims.	— luteum Retz.
— nigrifolia Desf.	— Solanum angustifolium Schrad.	— spathulatum Horn.	— luteum Retz.
— Schizanthus Waltheri Sims.	— Solanum angustifolium Schrad.	— tabulaforme Hae.	— luteum Retz.
— Scholertia alba Ledeb.	— Solanum angustifolium Schrad.	— tortuosum Ait.	— luteum Retz.
— bacifera C. A. M.	— Solanum angustifolium Schrad.	— Smithii Sims.	— luteum Retz.
— corniculata C. A. Meyer.	— Solanum angustifolium Schrad.	— spathulatum Horn.	— luteum Retz.

Caryopteris mastachanthus Schum.	Corchoropsis ornata Sieb. et Zucc.	Euphorbia spicatantha L.
Cassia marylandica L.	Corchorus leucostylis Hort.	Forsythia viridis DC.
Ceanothus azureus Desf.	Coreopsis grandiflora Nutt.	Fernia communis L.
— americanus L.	Coriandrum sativum L.	Ficifolia floridana Moq.
Cedrela sinensis Juss.	Coronilla montana Scop.	Glycyrrhiza echinata L.
Cenchrus tribuloides L.	Crataegus crus-galli L.	Glandelia lanceolata Nutt.
Centauria babylonica L.	Crotalaria incana L.	Guizotia chlorifera DC.
— macrocephala Murr.	Cucumis prophetarum L.	Gunnera scabra Retz.
— salmantica L.	Cucurbita aurantia Willd.	Guthriea viscosa Regel.
Ceranthus calceolatus Desf.	— melanosperma Braun.	Hebecladia dentata L.
Cephalaria katarica Ledeb.	Dalea lagopus Willd.	Helianthemum apenninum (L.) Mill.
Cestrum vespertinum L.	Darlingtonia brachyloba DC.	— var. roseum Grosser
Chenopodium anthelminticum L.	Dracopis latula L.	— nummularium (L.) Mill.
— Botrys L.	Delphinium Staphisagria	181 Tuberaria guttata (L.) Fourr.
— rubrum L.	Desmodium marylandicum Boott.	
Chilophyllum globosum DC.	Dianthus monosperma Boiss.	
Chrysanthemum carinatum Schlegel.	— americana Boiss.	
— corymbosum L.	Digitalis lutea L.	
— lacustris Brot.	Dioscorea divaricata L.	
Cirsium eriophorum Scop.	Dolichos labialis L.	
Cissus Roylei Hort.	Echinops spinosus L.	
Cistus salicifolius L.	Enelia calva Gray.	
Cladium mariscus R. Br.	Ephedra acrostichum Griseb.	
Clematis paniculata Ehrh.	Erigeron speciosus DC.	
— recta L.	Eryngium planum L.	
Clerodendron Bungei Hand.	Euphorbia cistagalli L.	
Clinopodium vulgare L.	— britannica L.	
Collomia grandiflora Bongl.	— foeniculatum L.	
Convolvulus scammonia L.	— Preslii Juss.	

167 Chenopodium capitatum (L.) Aschers.	
168 — ficifolium Sm.	
169 — foliosum Aschers.	
170 — opulifolium Schrad.	
171 — schradecianum Schult.	
172 — urticum L.	
173 Corispermum leptopterum (Asch.) Iljin	
174 Kochia scoparia (L.) Schrad.	
<u>Cistaceae</u>	
175 Cistus crispus L.	
176 — hirsutus Lam.	
177 — monspeliensis L.	
178 Helianthemum apenninum (L.) Mill.	
179 — var. roseum Grosser	
180 — nummularium (L.) Mill.	
181 Tuberaria guttata (L.) Fourr.	
<u>Cneoraceae</u>	
182 Cneorum tricoccon L.	
<u>Commelinaceae</u>	
183 Anemone papuanum Warburg	
184 Commelina communis L.	
185 — tuberosa L.	
186 Palisota mannii C.B. Clarke	
187 Tinantia erecta (Jacq.) Schlechtend.	
+ 188 Tradescantia crassifolia Cav.	
<u>Compositae (Asteraceae)</u>	
189 Achillea ageratifolia (Sibth. & Smith) Boiss.	
190 — macrophylla L.	
! 191 — millefolium L. (NG)	
192 Ammobium alatum R.Br.	
193 Antennaria howellii Greene	
194 — plantaginea R.Br.	
195 Anthemis arvensis L.	
196 — austriaca Jacq.	
197 — cotula L.	
! 198 — tinctoria L. (NG)	

Göttingen, Germany, 1841

Bordeaux, France, 1921

Göttingen, Germany, 1970

Seedlists

Bestell- Nr. / Order Nr.	Name	Herkunft u. Art der Diasporen / origin and type of diaspores	IPEN	Aufsammlungs- daten / collecting data
Asteraceae (cont.)				
30	<i>Centaurea jacea</i> L.	WS	AT-0-WU-0030521	42
31	-- <i>jacea</i> subsp. <i>angustifolia</i> (DC.) Gremli	WS	AT-0-WU-0031475	117
32	-- <i>pseudophrygia</i> C. A. Mey.	WS	AT-0-WU-0027271	125
33	-- <i>scabiosa</i> L.	WS	AT-0-WU-0030447	37
34	-- <i>scabiosa</i> L.	WS	AT-0-WU-0031441	95
35	<i>Oxysium cinum</i> (L.) All.	WS	AT-0-WU-0031471	121
36	-- <i>eriodorum</i> (L.) Scop.	WKS	AT-0-WU-0007975	125
37	-- <i>eristhales</i> (Jacq.) Scop.	WS	AT-0-WU-0031498	114
38	-- <i>pannonicum</i> (L. f.) Link	WS	AT-0-WU-0030486	37
39	-- <i>spinossissimum</i> (L.) Scop.	WS	AT-0-WU-0030577	53
40	<i>Crepis pyrenaica</i> (L.) Greuter	WS	AT-0-WU-0030527	45
41	-- <i>pyrenaica</i> (L.) Greuter	WS	AT-0-WU-0031494	114
42	<i>Cyanus segetum</i> Hill	WS	AT-0-WU-0031457	99
43	<i>Echinops ritro</i> subsp. <i>ruthenicus</i> (M. Bieb.) Nyman	WS	AT-0-WU-0030642	70
44	-- <i>ritro</i> subsp. <i>ruthenicus</i> (M. Bieb.) Nyman	WKS	AT-0-WU-0031014	125
45	<i>Erigeron acris</i> L.	WS	AT-0-WU-0030619	65
46	-- <i>glabellus</i> Nutt.	KS	XX-0-WU-0022433	125
47	<i>Eupatorium cannabinum</i> L.	KS	XX-0-WU-0025821	125
48	<i>Filago anensis</i> L.	WS	AT-0-WU-0031534	125
49	-- <i>vulgaris</i> Lam.	WS	AT-0-WU-0031367	111
50	<i>Galatella cana</i> (Waldst. & Kit.) Nees	WS	AT-0-WU-0030647	75
51	-- <i>linosyris</i> (L.) Rchb. f.	WS	AT-0-WU-0030644	72
52	<i>Gnaphalium sylvaticum</i> L.	WS	AT-0-WU-0029392	16
53	<i>Helenium aromaticum</i> (Hook.) L. H. Bailey	KS	XX-0-WU-0031245	125
54	<i>Hieracium prenanthoides</i> Vill.	KS	XX-0-WU-0030679	125
55	-- <i>umbellatum</i> L.	KS	XX-0-WU-0002204	77
56	-- <i>umbellatum</i> L.	WKS	AT-0-WU-0008873	125
57	-- <i>velutinum</i> Hegetschw.	WKS	CH-0-WU-0009851	125
58	-- <i>villosum</i> Jacq.	WKS	FR-0-WU-0008195	125
59	<i>Homogyne sylvestris</i> Cass.	WS	AT-0-WU-0029414	18
60	<i>Inula ensifolia</i> L.	WS	AT-0-WU-0030639	70
61	-- <i>oculus-christi</i> L.	WS	AT-0-WU-0030435	36
62	<i>Jurinea mollis</i> (L.) Rchb.	WS	AT-0-WU-0031346	91
63	<i>Klasea quinquefolia</i> (Willd.) Greuter & Wagenitz	KS	XX-0-WU-00319038	125
64	<i>Leontodon incanus</i> (L.) Schrank	WS	AT-0-WU-0027771	1
65	-- <i>pyrenaicus</i> subsp. <i>helveticus</i> (Mérat) Finch & P. D. Sell	WKS	FR-0-WU-0009648	125
66	<i>Leucanthemopsis alpina</i> (L.) Heywood	WS	AT-0-WU-0030566	53
67	<i>Liatris cylindracea</i> Michx.	WKS	NZ-0-MO-2015-0204	125
68	<i>Onopordum acanthium</i> L.	WS	AT-0-WU-0031459	96
69	<i>Prenanthes purpurea</i> L.	WKS	CH-0-WU-0007600	125
70	-- <i>purpurea</i> L.	WS	AT-0-WU-0031502	114
71	<i>Pseudognaphalium luteoalbum</i> (L.) Hilliard & B. L. Burt	KS	XX-0-WU-0031535	125

PARS I

SEMINA PLANTARUM IN LOCO NATALI LECTA

MAGNOLIOPHYTA MAGNOLIOPSIDA

ADOXACEAE

1. *Viburnum opulus* L. – Vitosha Mts, near Zlatni mostove locality, 19. 09. 2020.
2. *Sambucus racemosa* L. – Rila Mt, near Borovetz resort, 13. 08. 2020.

ANACARDIACEAE

3. *Pistacia terebinthus* L. – Managed reserve Izgoryaloto Gijune, near Krichim town, 12. 09. 2020.

ASTERACEAE

4. *Inula ensifolia* L. – Nature Park Zlatni pyasatsi, near Varna town, 28. 08. 2020.
5. *Jacobaea paludosa* (L.) G. Gaertn., B. Mey. & Scherb. (syn. *Senecio paludosus* L.) – Rodopi Mt, Tsigov chark locality, 16. 10. 2019.
6. *Tanacetum corymbosum* (L.) Sch. Bip. – Nature Park Zlatni pyasatsi, near Varna town, 28. 08. 2020.

BRASSICACEAE

7. *Alliaria petiolata* (M. Bieb.) Cavara & Grande – near Sotrya village, Stara Planina Mt., 05. 09. 2020.
8. *Lunaria annua* L. – Managed reserve Izgoryaloto Gijune, near Krichim town, 12. 09. 2020.

CAMPANULACEAE

9. *Campanula sibirica* L. – Nature Park Zlatni pyasatsi, near Varna town, 28. 08. 2020.

CAPRIFOLIACEAE

10. *Scabiosa argentea* L. – Nature Park Zlatni pyasatsi, near Varna town, 28. 08. 2020.

CARYOPHYLLACEAE

11. *Silene atropurpurea* (Griseb.) Greuter & Burdet (syn. *Viscaria atropurpurea* Griseb.), – near Dolen village, Western Rhodopes Mt., 30. 08. 2020.
12. *Silene atropurpurea* (Griseb.) Greuter & Burdet (syn. *Viscaria atropurpurea* Griseb.) – Samokov plane, near Samokov town, 13. 08. 2020.

FABACEAE

13. *Colutea arborescens* L. – Managed reserve Izgoryaloto Gijune, near Krichim town, 12. 09. 2020.

14. *Hippecrepis emerus* (L.) Lassen (syn. *Coronilla emerus* L.) – Managed reserve Izgoryaloto Gijune, near Krichim town, 12. 09. 2020.

15. *Trifolium angustifolium* L. – Managed reserve Izgoryaloto Gijune, near Krichim town, 12. 09. 2020.

LAMIACEAE

16. *Clinopodium vulgare* L. – Managed reserve Izgoryaloto Gijune, near Krichim town, 12. 09. 2020.
17. *Salvia sclarea* L. – River Struma valley, near Simidi Town 30. 08. 2020

OLEACEAE

18. *Fraxinus ornus* L. – Managed reserve Izgoryaloto Gijune, near Krichim town, 12. 09. 2020.
19. *Syringa vulgaris* L. – Managed reserve Izgoryaloto Gijune, near Krichim town, 12. 09. 2020.

PLANTAGINACEAE

20. *Digitalis lanata* Ehrh. - Managed reserve Izgoryaloto Gijune, near Krichim town, 12. 09. 2020

PRIMULACEAE

21. *Primula veris* L. – near Dolen village, Rhodopes Mt., 30. 08. 2020.

RHAMNACEAE

22. *Paliurus spina-christi* Mill. - Managed reserve Izgoryaloto Gijune, near Krichim town, 12. 09. 2020

SOLANACEAE

23. *Physalis alkekengi* L. - near Dolen village, Rhodopes Mt., 30. 08. 2020.

LILIOPSIDA

ALLIACEAE

24. *Allium flavum* L. – near Luki village, Central Rhodopes Mt. 30. 08. 2020.

IRIDACEAE

25. *Iris sibirica* L. – Vitosha Mt., 31. 08. 2020.

POACEAE

26. *Danthonia alpina* Vest (syn. *Danthonia calycina* Roem. & Schult.) – Vitosha Mt., 31. 08. 2020.

XANTHORRHOACEAE

27. *Asphodeline lutea* (L.) Rchb. – Nature Reserve Kaliakra, 30. 10. 2020.
28. *Asphodelus albus* Mill. – Vitosha Mt., 31. 08. 2020.

Collectors:

Antoaneta Petrova, PhD
Diana Venkova

Nr	Familie	Gattung	Species (Subspecies, Varietät)	Autor	IPEN
76	Asteraceae	Liatris	elegans	Michx.	XX-0-LI-7086
77	Asteraceae	Liatris	punctata	Hook.	XX-0-LI-10164
78	Asteraceae	Liatris	spicata	Willd.	XX-0-LI-7094
79	Asteraceae	Liatris	spicata	Willd.	XX-0-LI-10268
80	Asteraceae	Onopordum	algeriense	Pomel	XX-0-LI-6994
81	Asteraceae	Onopordum	bracteatum	Boiss. & Heldr.	XX-0-STGAL-130/1981
82	Asteraceae	Osteospermum	hyoseroides	(DC.) Norl.	XX-0-LI-14339
83	Asteraceae	Pilosella	aurantiaca	(L.) F.W. Schultz & Sch.Bip.	XX-0-LI-11633
84	Asteraceae	Scorzonera	suberosa ssp. carlensis	(Boiss.) D.F. Chamb.	XX-0-LI-13055
85	Asteraceae	Senecio	alpinus	Scop.	XX-0-LI-7426
86	Asteraceae	Silphium	perfoliatum	L.	XX-0-LI-10168
87	Asteraceae	Solidago	virgaurea	L.	XX-0-LI-14222
88	Asteraceae	Stockesia	laevis	Greene	XX-0-LI-7111
89	Asteraceae	Tanacetum	corymbosum	(L.) Sch.Bip.	IT-0-SIENA-A-10
90	Asteraceae	Telekia	speciosa	(Schreb.) Baumg.	XX-0-LI-8899A
91	Berberidaceae	Berberis	koreana	PALIB.	XX-0-LI-7898
92	Bignoniaceae	Incarvillea	delavayi	Bureau et Franch	XX-0-LI-6808
93	Boraginaceae	Echium	vulgare	L.	XX-0-LI-8987
94	Boraginaceae	Onosma	ablorosea	Fisch.et C.A. Mey	XX-0-LI-6693
95	Brassicaceae	Alyssum	altaicum	C.A.Mey.	XX-0-LI-6619
96	Brassicaceae	Alyssum	ovirense	A.Kern.	XX-0-LI-7290
97	Brassicaceae	Alyssum	wulfenianum	Bernh.	XX-0-LI-6461
98	Brassicaceae	Arabis	blepharophylla	Hook. et Arn.	XX-0-LI-6398

Vienna, Austria, 2020

Sofia, Bulgaria, 2020

Linz, Austria, 2020

Use case : Seedlists

Bestell- Nr. / Order Nr.	Name	Herkunft u. Art der Diasporen / origin and type of diaspores	IPEN	Aufsammlungs- daten / collecting data
Asteraceae (cont.)				
30	<i>Centaurea jacea</i> L.	WS	AT-0-WU-0030521	42
31	-- <i>jacea</i> subsp. <i>angustifolia</i> (DC.) Gremli	WS	AT-0-WU-0031475	117
32	-- <i>pseudophrygia</i> C. A. Mey.	WS	AT-0-WU-0027271	125
33	-- <i>scabiosa</i> L.	WS	AT-0-WU-0030447	37
34	-- <i>scabiosa</i> L.	WS	AT-0-WU-0031441	95
35	<i>Cirsium canum</i> (L.) All.	WS	AT-0-WU-0031471	121
36	-- <i>eriphorum</i> (L.) Scop.	WKS	AT-0-WU-0007975	125
37	-- <i>eristhales</i> (Jacq.) Scop.	WS	AT-0-WU-0031498	114
38	-- <i>pannonicum</i> (L. f.) Link	WS	AT-0-WU-0030468	37
39	-- <i>spinosissimum</i> (L.) Scop.	WS	AT-0-WU-0030577	53
40	<i>Crepis pyrenaica</i> (L.) Greuter	WS	AT-0-WU-0030527	45
41	-- <i>pyrenaica</i> (L.) Greuter	WS	AT-0-WU-0031494	114
42	<i>Cyanus segetum</i> Hill	WS	AT-0-WU-0031457	99
43	<i>Echinops ritro</i> subsp. <i>ruthenicus</i> (M. Bieb.) Nyman	WS	AT-0-WU-0030642	70
44	-- <i>ritro</i> subsp. <i>ruthenicus</i> (M. Bieb.) Nyman	WKS	AT-0-WU-0031014	125
45	<i>Erigeron acris</i> L.	WS	AT-0-WU-0030619	65
46	-- <i>glaberrimus</i> L.	KS	XX-0-WU-0022433	125
47	<i>Eupatorium cannabinum</i> L.	KS	XX-0-WU-0025821	125
48	<i>Filago anensis</i> L.	WS	AT-0-WU-0031534	125
49	-- <i>vulgaris</i> Lam.	WS	AT-0-WU-0031367	111
50	<i>Galatella cana</i> (Waldst. & Kit.) Nees	WS	AT-0-WU-0030647	75
51	-- <i>linosyris</i> (L.) Rchb. f.	WS	AT-0-WU-0030644	72
52	<i>Gnaphalium sylvaticum</i> L.	WS	AT-0-WU-0029392	16
53	<i>Helenium aromaticum</i> (Hook.) L. H. Bailey	KS	XX-0-WU-0031245	125
54	<i>Heracium prenanthoides</i> Vill.	KS	XX-0-WU-0030679	125
55	-- <i>umbellatum</i> L.	KS	XX-0-WU-0002204	77
56	-- <i>umbellatum</i> L.	WKS	AT-0-WU-0008873	125
57	-- <i>velutinum</i> Hegetschw.	WKS	CH-0-WU-0009851	125
58	-- <i>villosum</i> Jacq.	WKS	FR-0-WU-0008195	125
59	<i>Homogyne sylvestris</i> Cass.	WS	AT-0-WU-0029414	18
60	<i>Inula ensifolia</i> L.	WS	AT-0-WU-0030639	70
61	-- <i>oculus-christi</i> L.	WS	AT-0-WU-0030435	36
62	<i>Jurinea mollis</i> (L.) Rchb.	WS	AT-0-WU-0031346	91
63	<i>Klasea quinquefolia</i> (Willd.) Greuter & Wagenitz	KS	XX-0-WU-0019038	125
64	<i>Leontodon incanus</i> (L.) Schrank	WS	AT-0-WU-0027771	1
65	-- <i>pyrenaicus</i> subsp. <i>helveticus</i> (Mérat) Finch & P. D. Sell	WKS	FR-0-WU-0009648	125
66	<i>Leucanthemopsis alpina</i> (L.) Heywood	WS	AT-0-WU-0030566	53
67	<i>Liatris cylindracea</i> Michx.	WKS	NZ-0-HO-2015-0204	125
68	<i>Onopordum acanthium</i> L.	WS	AT-0-WU-0031459	96
69	<i>Prenanthes purpurea</i> L.	WKS	CH-0-WU-0007600	125
70	-- <i>purpurea</i> L.	WS	AT-0-WU-0031502	114
71	<i>Pseudognaphalium luteoalbum</i> (L.) Hilliard & B. L. Burt	KS	XX-0-WU-0031535	125

- 30 *Centaurea jacea* L.
- 31 -- *jacea* subsp. *angustifolia* (DC.) Gremli
- 32 -- *pseudophrygia* C. A. Mey.
- 33 -- *scabiosa* L.
- 34 -- *scabiosa* L.
- 35 *Cirsium canum* (L.) All.
- 36 -- *eriphorum* (L.) Scop.
- 37 -- *eristhales* (Jacq.) Scop.
- 38 -- *pannonicum* (L. f.) Link
- 39 -- *spinosissimum* (L.) Scop.
- 40 *Crepis pyrenaica* (L.) Greuter
- 41 -- *pyrenaica* (L.) Greuter
- 42 *Cyanus segetum* Hill
- 43 *Echinops ritro* subsp. *ruthenicus* (M. Bieb.) Nyman
- 44 -- *ritro* subsp. *ruthenicus* (M. Bieb.) Nyman
- 45 *Erigeron acris* L.

Vienna, Austria, 2020

Seedlists

Research Engineering task

- Extract plant species names from seedlists

Scientific goal of the project

- Create database
- Analysis:
 - Curational policies
 - Effects of climate change

Prompt

You are an expert botanist with deep knowledge of Latin names in binomial nomenclature of plant species.

You have knowledge of plant family names, subspecies, varieties and forms.

The species name in binomial nomenclature consists of two parts: the genus name starting with a capital letter, followed by the epithet name which normally starts with a small letter.

This is often followed by a subspecies name (starting with "subsp. " or "ssp. "), or variety name (starting with "var."), or form (starting with "f.") and the author name(s). Sometimes this is followed by a synonym (starting with a syn.).

Some species names are also followed by a cultivar name (almost always within single quotation marks or with the abbreviation "CV" if the cultivar has no formal name).

The author name is often simply "L.", for "Linnaeus". For other authors it is often an abbreviation describing the author name.

Thus the species name consists of the genus, epithet, subspecies name (if present), variety name (if present), form name (if present), cultivar name (if present), author name(s) (if present) and synonym (if present).

You are given the text from a page from a seedlist of plants from a botanical garden.

Determine if there are any Latin names of plant species present in this text.

Find each and every plant species entry present in the text.

Do not skip or miss any entry.

If any plant species names are present in this text, find, for each plant species present in the text, the species name in Latin, consisting of the genus, epithet, subspecies name (if present), variety name (if present), form name (if present), cultivar name (if present), author name(s) (if present), and synonym (if present).

Return the result for each plant species in valid JSON object format, with the single key "species".

Provide the complete results as a list of valid JSON objects.

If there is no information about plant species on this page, print an empty list, like this: []

Print only the complete list of valid JSON objects, and no other code, text, or explanation.

Set
AI persona

Provide
background
information

Describe
task

Specify
output format

Temperature parameter

0



2



Less random
More consistent
Less creative

More random
Less consistent
More creative

Temperature = 0 : Maximises both accuracy and consistency

Output

Asteraceae (cont.)

- 30 *Centaurea jacea* L.
31 -- *jacea* subsp. *angustifolia* (DC.) Greml
32 -- *pseudophrygia* C. A. Mey.
33 -- *scabiosa* L.
34 -- *scabiosa* L.
35 *Cirsium canum* (L.) All.
36 -- *eriphorum* (L.) Scop.
37 -- *erisithales* (Jacq.) Scop.
38 -- *pannonicum* (L. f.) Link
39 -- *spinosissimum* (L.) Scop.
40 *Crepis pyrenaica* (L.) Greuter
41 -- *pyrenaica* (L.) Greuter
42 *Cyanus segetum* Hill
43 *Echinops ritro* subsp. *ruthenicus* (M. Bieb.) Nyman
44 -- *ritro* subsp. *ruthenicus* (M. Bieb.) Nyman
45 *Erigeron acris* L.

Input

- | | |
|----|---|
| 0 | <i>Centaurea jacea</i> L. |
| 1 | <i>Centaurea jacea</i> subsp. <i>angustifolia</i> (DC.) Greml |
| 2 | <i>Centaurea pseudophrygia</i> C. A. Mey. |
| 3 | <i>Centaurea scabiosa</i> L. |
| 4 | <i>Centaurea scabiosa</i> L. |
| 5 | <i>Cirsium canum</i> (L.) All. |
| 6 | <i>Cirsium eriphorum</i> (L.) Scop. |
| 7 | <i>Cirsium erisithales</i> (Jacq.) Scop. |
| 8 | <i>Cirsium pannonicum</i> (L. f.) Link |
| 9 | <i>Cirsium spinosissimum</i> (L.) Scop. |
| 10 | <i>Crepis pyrenaica</i> (L.) Greuter |
| 11 | <i>Crepis pyrenaica</i> (L.) Greuter |
| 12 | <i>Cyanus segetum</i> Hill |
| 13 | <i>Echinops ritro</i> subsp. <i>ruthenicus</i> (M. Bieb.) Nyman |
| 14 | <i>Echinops ritro</i> subsp. <i>ruthenicus</i> (M. Bieb.) Nyman |
| 15 | <i>Erigeron acris</i> L. |

Output

Seedlists: OCR error correction

Original typewritten text

167 Chenopodium capitatum (L.) Aschers.
168 - ficifolium Sm.
169 - foliosum Aschers.
170 - opulifolium Schrad.
171 - schraderianum Schult.
172 - urbicum L.
173 Corispermum leptopterum (Asch.) Iljin
174 Kochia scoparia (L.) Schrad.

C i s t a c e a e
175 Cistus crispus L.
176 - hirsutus Lam.
177 - monspeliensis L.
178 Helianthemum apenninum (L.) Mill.
179 - - var. roseum Grosser
180 - nummularium (L.) Mill.
181 Tuberaria guttata (L.) Fourr.

C n e o r a c e a e
182 Cneorum tricoccon L.

C o m m e l i n a c e a e
183 Aneilema papuanum Warburg
184 Commelina communis L.
185 - tuberosa L.
186 Palisota mannii C.B. Clarke
187 Tinantia erecta (Jacq.) Schlechtend.
+ 188 Tradescantia crassifolia Cav.

C o m p o s i t a e (Asteraceae)
189 Achillea ageratifolia (Sibth. & Smith)
Boiss.
190 - macrophylla L.
! 191 - millefolium L. (NG)
192 Ammobium alatum R.Br.
193 Antennaria howellii Greene
194 - plantaginea R.Br.
195 Anthemis arvensis L.
196 - austriaca Jacq.
197 - cotula L.
! 198 - tinctoria L. (NG)

Extracted OCR

Chenopodium capitatum (L.) Aschers»
168 - * ficifolium Sm.
■ 169 - foliosum Asciesse.
170 ~ opulifolium Schaad»
171 - schraderianum Schult»
172 - urbicum L»
175 Corispermum leptopterum (Asch.) Iljin
17^ Kochia scoparia (L») Schrad»
Cista c e-a e
175 Cistus crispus L»'
176 - hirsutus Lam»
177 - monspeliensis L».,,
178 Heliantheum apenninum (L») Mill»
179 - - var» roseum Grossor
180 - nummularium (L») Mill»
181 Tuberaria guttata (L») Fourr»
C n e o r a c e a e
182 Cneorum tricoccon L»
C o m m e l i n a c e a _e
185 Aneilema papuanum Warburg
18^ Commelina communis L»
185 - tuberosa L»
186 Palisota mannii C»B»ClarkeN
187 Tinantia erecta (Jacq») Schlechtend»
+ 188 Tradescantia crassifolia Cav».
Compositae (Asteraceao)
189 Achillea ageratifolia (Sibth» & Smith)
Boiss »
190 - macrophylla L»
! 191 - millefolium L» (NG)
192 Ammobium alatum R»Br.
195 Antennaria howellii Greene
19^ - plantaginea R»Br»
195 Anthemis arvensis L»
196 - austriaca Jacq»
197 - cotula L»
! 198 - tinctoria L» (NG)

Claude 3 Opus output

0	Chenopodium capitatum (L.) Aschers.
1	Chenopodium ficifolium Sm.
2	Chenopodium foliosum Aschers.
3	Chenopodium opulifolium Schrad.
4	Chenopodium schraderianum Schult.
5	Chenopodium urbicum L.
6	Corispermum leptopterum (Asch.) Iljin
7	Kochia scoparia (L.) Schrad.
8	Cistus crispus L.
9	Cistus hirsutus Lam.
10	Cistus monspeliensis L.
11	Helianthemum apenninum (L.) Mill.
12	Helianthemum apenninum var. roseum Grosser
13	Helianthemum nummularium (L.) Mill.
14	Tuberaria guttata (L.) Fourr.
15	Cneorum tricoccon L.
16	Aneilema papuanum Warburg
17	Commelina communis L.
18	Commelina tuberosa L.
19	Palisota mannii C.B. Clarke
20	Tinantia erecta (Jacq.) Schlechtend.
21	Tradescantia crassifolia Cav.
22	Achillea ageratifolia (Sibth. & Smith) Boiss.
23	Achillea macrophylla L.
24	Achillea millefolium L.
25	Ammobium alatum R.Br.
26	Antennaria howellii Greene
27	Antennaria plantaginea R.Br.
28	Anthemis arvensis L.
29	Anthemis austriaca Jacq.
30	Anthemis cotula L.
31	Anthemis tinctoria L.

Use case: Health Technology Assessment (HTA) documents

NICE National Institute for Health and Care Excellence

HAS
HAUTE AUTORITÉ DE SANTÉ



Zorginstituut Nederland

Ivabradine for treating chronic heart failure (TA267)

1 Recommendations

- 1.1 Ivabradine is recommended as an option for treating chronic heart failure for people:
- with New York Heart Association (NYHA) class II to IV stable chronic heart failure with systolic dysfunction and
 - who are in sinus rhythm with a heart rate of 75 beats per minute (bpm) or more and
 - who are given ivabradine in combination with standard therapy including beta-blocker therapy, angiotensin-converting enzyme (ACE) inhibitors and aldosterone antagonists, or when beta-blocker therapy is contraindicated or not tolerated and
 - with a left ventricular ejection fraction of 35% or less.
- 1.2 Ivabradine should only be initiated after a stabilisation period of 4 weeks on optimised standard therapy with ACE inhibitors, beta-blockers and aldosterone antagonists.

1. CARACTERISTIQUES DU MEDICAMENT

1.1. Principe actif
Ivabradine

1.2. Indications

« Traitement de la maladie coronaire :

Traitement symptomatique de l'angor stable chronique chez l'adulte coronarien en rythme sinusal. L'ivabradine est indiquée :

- chez les adultes présentant une intolérance ou une contre-indication aux bêtabloquants,
- ou en association aux bêtabloquants chez des patients insuffisamment contrôlés par une dose optimale de bêtabloquants, et dont la fréquence cardiaque reste supérieure à 60 bpm ».

Traitement de l'insuffisance cardiaque : nouvelle indication

L'ivabradine est indiquée dans le traitement de l'insuffisance cardiaque chronique de classe NYHA II à IV avec dysfonction systolique, chez les patients en rythme sinusal et dont la fréquence cardiaque est supérieure ou égale à 75 bpm, en association au traitement standard comprenant les bêtabloquants, ou en cas de contre-indication ou d'intolérance aux bêtabloquants (cf Pharmacodynamie) ».

1.3. Posologie

« Pour les différentes doses thérapeutiques, l'ivabradine est disponible sous forme de comprimés pelliculés dosés à 5 mg et 7,5 mg.

Traitement de l'insuffisance cardiaque chronique :

Le traitement doit être initié uniquement chez les patients atteints d'insuffisance cardiaque stable. Il est recommandé que le médecin ait une expérience de la prise en charge de l'insuffisance cardiaque chronique.

La posologie initiale habituellement recommandée est de 5 mg d'ivabradine deux fois par jour. Après deux semaines de traitement, la dose peut être augmentée à 7,5 mg deux fois par jour si la fréquence cardiaque de repos reste de façon persistante au dessus de 60 bpm, ou diminuée à 2,5 mg deux fois par jour (soit un demi comprimé dosé à 5 mg deux fois par jour) si la fréquence cardiaque de repos est de façon persistante en dessous de 50 bpm ou en cas de symptômes liés à la bradycardie tels que sensations vertigineuses, fatigue ou hypotension. Si la fréquence cardiaque est comprise entre 50 et 60 bpm, la posologie de 5 mg deux fois par jour peut être maintenue.

Si, durant le traitement, la fréquence cardiaque de repos descend de façon persistante en dessous de 50 bpm ou en cas de symptômes liés à la bradycardie, la dose doit être diminuée à la posologie inférieure chez les patients recevant 7,5 mg ou 5 mg deux fois par jour.

Farmacotherapeutisch rapport ivabradine (Procoralan®) bij de indicatie chronisch hartfalen

De Commissie Geneesmiddelen (CG) heeft een farmacotherapeutisch rapport vastgesteld voor het geneesmiddel ivabradine (Procoralan®) filmomhulde tabletten. Voor de bepaling van de therapeutische waarde is vergeleken met placebo. Hierbij is zij tot onderstaande conclusies gekomen.

- Ivabradine heeft een therapeutische meerwaarde ten opzichte van placebo bij patiënten met chronisch hartfalen NYHA klasse II-IV in sinusritme met systolische disfunctie met rusthartfrequentie ≥ 75 spm wanneer toegevoegd aan een individueel, optimaal en richtlijnconform ingestelde medicamenteuze behandeling die in principe een ACE-remmer, diureticum, bètablokker en aldosteronantagonist omvat tenzij de patiënt hiervoor niet is aangewezen, een contra-indicatie heeft of deze niet verdraagt.

Ivabradine (Procoralan®), filmomhulde tablet

Geregistreerde indicatie. « (...) chronisch hartfalen NYHA-klasse II-IV met systolische dysfunctie, bij patiënten in sinusritme en bij wie de hartfrequentie ≥ 75 spm is, in combinatie met standaardbehandeling, inclusief behandeling met bètablokkers, of wanneer behandeling met bètablokkers is gecontra-indiceerd of niet wordt getolereerd. »

Dosering. 5 mg 2dd. Na een behandelingsduur van twee weken kan de dosis worden verhoogd naar 7,5 mg 2dd wanneer de hartfrequentie in rust voortdurend boven de 60 spm is of worden verlaagd tot 2dd 2,5 mg wanneer de hartfrequentie in rust voortduren onder de 50 slagen per minuut spm is.

Werkingsmechanisme. Ivabradine werkt via selectieve en specifieke remming van de cardiale pacemaker I_f -stroom, welke de spontane diastolische depolarisatie in de sinusknop en de hartfrequentie bepaalt. De cardiale effecten zijn specifiek voor de sinusknop zonder effecten op intra-atriale, atrioventriculaire of intraventriculaire geleidingstijden, evenmin als op myocardiale contractiliteit of ventriculaire repolarisatie.

HTA documents: Attributes to extract

Data point	Explanation
HTA ID	Name of HTA organisation performing the assessment
Assessment type	Is this the first assessment, a reassessment, or an indication broadening?
Internal identifier	Code or label identifying the document
INN	International non-proprietary name of assessed drug
Brand name	Brand name of assessed drug
Assessment date	When was the assessment finalised?
Indication	Medical condition for which the drug is assessed
Final recommendation	What is the final recommendation for this drug-indication combination?
Comparator	Drug(s) with which the performance of the assessed drug is compared
Relative effectiveness assessment outcome	Outcome of the relative effectiveness assessment for this drug-indication combination
Cost-effectiveness assessment outcome	Outcome of the cost-effectiveness assessment for this drug-indication combination
Budget impact outcome	Budget impact of adoption of the drug
Managed entry agreements	Was any OECD-defined managed entry agreement proposed? If so, which class?
Clinical restrictions	Clinical restrictions stated in the recommendation

HTA documents

Research Engineering task

- Extract certain data points of interest from HTA documents

Scientific goal of the project

- Create Open Science database that can be used by other researchers, policymakers, etc.

Use case : Kickstarter

KICKSTARTER

Q Search projects, creators, and categories

Start a project


Art Comics Crafts Dance Design Fashion Film Food Games Journalism Music Photography Publishing Technology Theater Discover

Bring a creative project to life.

ON KICKSTARTER:

264,804 projects funded	\$8,227,855,974 towards creative work	97,266,101 pledges
----------------------------	--	-----------------------

FEATURED PROJECT




Swedish Machines – New Simon Stålenhag Book
Free League
2 days left • 1028% funded


A new art book from a critically acclaimed illustrator and author, set in a beautiful dystopian world in which two young men set out to explore a mysterious forbidden zone together.

Illustration Stockholm, Sweden


RECOMMENDED FOR YOU




Valuchi Expedition...
Valuchi
24 days left • 246%...



Tir na nÓg Reprint...
Grand Gamers Guild
3 days left • 6126%...



THE DISCOVERIE ...
Black Letter Press
5 days left • 415%...



Bulletins at The En...
Russell Nohetty
9 days left • 133%...

< 1 2 3 >

Kickstarter: NAICS codes

← ↻ <https://www.census.gov/naics/?58967?yearbck=2017>

An official website of the United States government

Census
BUREAU OF ECONOMIC ANALYSIS

North American Industry Classification System

Main History Development Partners Federal Register Notices FAQs Reference Files ▾ NAPCS

NAICS Search

Enter keyword or 2-6 digit code

2022 NAICS Search Go

Enter keyword or 2-6 digit code

2017 NAICS Search Go

Enter keyword or 2-6 digit code

2012 NAICS Search Go

2017 NAICS

The following table provides detailed information on the structure of NAICS.

Sector	Definition
11	Agriculture, Forestry, Fishing and Hunting
21	Mining, Quarrying, and Oil and Gas Extraction
22	Utilities
23	Construction
31-33	Manufacturing
42	Wholesale Trade
44-45	Retail Trade
48-49	Transportation and Warehousing
51	Information
52	Finance and Insurance
53	Real Estate and Rental and Leasing
54	Professional, Scientific, and Technical Services
55	Management of Companies and Enterprises
56	Administrative and Support and Waste Management and Remediation Services
61	Educational Services
62	Health Care and Social Assistance
71	Arts, Entertainment, and Recreation
72	Accommodation and Food Services
81	Other Services (except Public Administration)
92	Public Administration

Sector 51--Information^T

511 Publishing Industries (except Internet)^T

5111 Newspaper, Periodical, Book, and Directory Publishers^T

51111 Newspaper Publishers^T

511110 Newspaper Publishers

51112 Periodical Publishers^T

511120 Periodical Publishers

51113 Book Publishers^T

511130 Book Publishers

51114 Directory and Mailing List Publishers^T

511140 Directory and Mailing List Publishers

51119 Other Publishers^T

511191 Greeting Card Publishers

511199 All Other Publishers

5112 Software Publishers^T

51121 Software Publishers^T

511210 Software Publishers

Industry code

Number of 4-digit industry codes: 311

Kickstarter

Research Engineering task

- Assign a NAICS code to every Kickstarter project

Scientific goal of the project

- Assess whether crowdfunding fosters innovation and economic development

Results (with Claude 3 Opus API)

Project	Results
Seedlists	<p>4 sample pages from seedlists, variety of different formats, total 125 plant species names:</p> <ul style="list-style-type: none">• All species names extracted, and extracted correctly. Consistent over 3 runs• Even some OCR errors corrected for scanned seedlists!
HTA documents	<p>HTA documents about one drug-indication combination, in English, Dutch, French:</p> <ul style="list-style-type: none">• 11/14 attributes extracted correctly. Consistent over 3 runs.• 2 attributes not lexically identical over the 3 runs, but semantically consistent• For one (HTA organization name), a subcommittee name instead of the parent French organization name was extracted.
Kickstarter	<p>Sample of 540 representative projects, assigned to six human raters in a staggered manner. Highest fraction of industry codes that matched between:</p> <ul style="list-style-type: none">• Generative AI and a (single) human rater: 53%,• Two human raters: 60%

Insights derived

When is generative AI an appropriate tool?

- Amount of data to be processed is large
- No simple, rule-based method is available
- Results are of sufficiently high quality for the research purpose
- Good tool for tasks that are easy for a human to perform, but difficult for a computer
 - Heterogeneous, unstructured data

Maximising accuracy and consistency

- **Temperature:**
 - Set to 0 to maximise accuracy and consistency
- **Prompt:**
 - Well-posed, unambiguous, clear
 - **Prompt engineering:** Iterate over many different versions of the prompt, with a representative input dataset
 - Prompt should be **general as well as specific:**
 - General wrt **data**
 - Specific wrt **AI's actions**

Advanced computing considerations

- Previously intractable natural language processing problems now solvable with generative AI
- This solution is still data- and compute-intensive, but the HPC part is abstracted away via the API to the proprietary model provider (OpenAI, Anthropic, etc.).

Advanced computing considerations: Proprietary, private, or sensitive data

- Enterprise cloud services:
 - OpenAI API through Microsoft's Azure OpenAI Service
 - Anthropic's Claude API through Amazon Bedrock
- HPC part still abstracted away by Azure/Amazon Bedrock.

Advanced computing considerations:

Reproducibility

- Proprietary model providers routinely deprecate older models
 - Results can no longer be reproduced.
 - Can be avoided by using open-weights models (e.g. Llama)
- Open-weights models can be downloaded and run locally, but this requires:
 - Significant computational resources
 - Technical know-how about what resources to use, and how
 - Raises barrier to entry for domain researchers

Thank you!
Questions?
m.mitra@uu.nl

IEEE International Conference on eScience 2024 paper:
<https://ieeexplore.ieee.org/document/10678704>



GitHub repository for data, prompts, code:
<https://github.com/UtrechtUniversity/generative-ai>



**Utrecht
University**

Sharing science,
shaping tomorrow