

Data and Democracy

The University of Chicago Data Science Institute worked to address a gap in election research: the absence of a comprehensive repository for historical U.S. election data. The Data & Democracy team developed an innovative approach to extract and compile election results from 1920 to 1996, leveraging advanced artificial intelligence to transform thousands of scanned paper records into structured, analyzable data.

The challenge involved developing prompts for the commercial API to accurately interpret complex election data from 51 PDFs containing over 3,000 pages of historical records. The team created a sophisticated pipeline that converted PDFs to images, processed them through the API's vision capabilities, and generated standardized CSV files with critical election metrics. Each record captured state, year, race type, district, candidate name, party affiliation, and vote totals.

Through prompt engineering and accuracy validation using edit distance and digit-level measurements, the project achieved exceptional accuracy rates: 100% for state, race type, district, and candidate party; 94% for candidate names; and 98% for vote totals. This digitized archive now provides researchers with unprecedented access to decades of previously inaccessible electoral history, advancing research on democratic institutions and governance.

Title of candidate	Democrat	Republican	American Inde- pendent	Socialist Labor	Socialist Workers	Others	Scat- tering	Total
Representatives:								
1st district		159, 282					147	159, 429
2d district	137, 616						9	137, 625
3d district	156, 703						136	156, 839
4th district	99, 977	93, 927				11, 141	2	205, 047
Sth district	92, 84/	110, 970	3,803				3	207, 623
7th district	135,057	/8, 381 64 357					12	100 562
8th district	142 470	04, 337			18 169		48	160 687
9th district	67 143	23 177			2 397	1 70 571	40	163, 288
10th district	07,140	161, 708			2, 557	70, 071	57	161, 765
11th district	154, 397	101,700					56	154, 453
12th district	117, 753	116, 547			•••••		4	234, 304
Total by districts	1, 243, 796	808, 349	3, 803		20, 566	81, 712	474	2, 158, 700
Senator	823, 278	1, 505, 932			41, 369		97	2, 370, 676
Presidential electors	1, 332, 540	1, 112, 078	2,877	129	10,600	2 190	342	2, 458, 756

¹ Independent Conservative Party. ² Includes 101 votes for Peace and Freedom Party, 46 votes for Communist Party, and 43 votes for Libertarian Party.

Figure 1: Example table from PDFs

Category Name	Accuracy		
State	100%		
Race Type	100%		
District	100%		
Candidate Name	94%		
Candidate Party	100%		
Vote Count	98%		

Figure 2: Accuracy Statistics