

# Lynn M. Cherny, Ph.D.

Email: [lynn@cherny.com](mailto:lynn@cherny.com)

[www.ghostweather.com](http://www.ghostweather.com)

Social: @arnicas ([twitter](#), [bluesky](#), [mastodon](#))

Lyon, France

## SUMMARY

- ◆ AI-savvy data science, especially NLP. I'm "full stack" for prototyping although I am not a front-end designer.
- ◆ Keywords: Data analysis, scraping, classification/entities/topics recognition, databases, data visualization, creative AI (especially generation), knowledge networks.
- ◆ Creative AI/narrative/DS news newsletter: [arnicas.substack.com](http://arnicas.substack.com) (6K+ subs)

## EMPLOYMENT HISTORY

### Consultant, Data Science

Oct 2019 - Now

### Ghostweather R&D, SARL (self-employed)

Remote from Lyon, France

- ◆ "Artist in Residence" at Google Arts & Culture for 4 years doing AI/creative apps – See below.
- ◆ AI/NLP applications and models – e.g., worked on [UK Dubai Pavilion AI poetry generation](#) with Kyle McDonald, toxic speech detection (see below at Spirit AI)
- ◆ Built early knowledge base for common sense relations for HiddenDoor.co, an AI game startup, plus word2vec and other methods for genre analysis
- ◆ Document tagging (entities and topics, using spaCy v3 and transformers) contract for Organized Crime and Corruption Reporting, web data scraping
- ◆ Tutorials for data visualization and analysis (e.g., FlowingData)

### ML Artist in Residence

Oct 2019 - Now

### Google Arts & Culture

Paris, France (working remotely)

- ◆ : [Where is Hopper](#) (ink-based game with early AI gen art), [Haiku Imagined](#) (curated and AI-illustrated with Google models), [Surfacing Women in Smithsonian History](#) (DS work on Smithsonian metadata for entity relations)
- ◆ Data-related analysis of partner data: [Smithsonian data analysis and BigQuery dashboards](#), James Beard recipes
- ◆ Creation of data sets, including use of SPARQL for wikipedia data retrieval, Gutenberg Books
- ◆ Presentations to partners on text visualization and generation, LLM use in digital collections analysis...

### Consultant, Data Science Lead

Oct 2018 – Oct 2019

### SpiritAI.com (and others)

Remote from Lyon, France

- ◆ NLP models for offensive chat classification (hate speech, harassment, etc). Used PyTorch, Tensorflow, SpaCy. Spark data analysis. Team lead and mentorship.
- ◆ Data analysis and visualization for creative AI applications, e.g. for narrative or text generation

### Associate Professor, Data Science

Sept 2016 – Sept 2018

### EMLyon Business School

Lyon, France

- ◆ Prepared and taught many full courses on data analysis, programming, AI, Python: *AI for Business* (a reading/demos course), *Text Mining with Python*, *Python Bootcamp*, *Python Pandas Data Analysis*, *NLP in Python*, *Business Analytics Tools* (Excel, SQL, Tableau), *Web Database Programming*. Teaching reviews > 4.
- ◆ Head of Data R&D Institute Development: Hosted training workshops for faculty and students, scheduled visiting research and practice talks, recruited and hired, and co-chaired OpenVis Conf for 6<sup>th</sup> year, hosted in Paris for the first time. Maintainer of web site and twitter account for Data Institute (@data\_emlyon).

- ◆ Research activities involving data scraping, collecting, social media and image analysis with machine learning, analysis of job ads (skills detection); invited conference talks.

**Visiting Knight Chair in Journalism (2015-2016)**

**University of Miami, School of Communication**

Sept 2015 – June 2016

- ◆ Academic Knight fellowship teaching Interactive Data Visualization (materials online: <http://arnicas.github.io/interactive-vis-course/>) and workshops on data analysis methods.
- ◆ Developing toolkit for teaching and visualizing text analysis with Bocoup.com

**Self-Employed Consultant for Data Analysis and Data Visualization Projects**

**Ghostweather Research & Design, LLC**

February 2007—2015

Boston Area

- ◆ Two years consulting for Cambridge financial startup **Kensho**: D3 and R data visualization, data analysis, NLP, interactive dashboard prototypes of entity recognition in news articles (GDELT and then our own).
- ◆ Three years consulting as data analyst and developer for **Brandeis University**: Bayesian models in R, data graphing in R and Python, Python stats/graphs/files for web site displaying research data, Javascript UI
- ◆ Sample NLP work: PDF doc parsing, entity extraction, topic classification/clustering (LDA, naïve bayes)
- ◆ Dashboard design: Identify key metrics, mockups of useful charts for data presentation, analysis of data
- ◆ D3.js and Python (matplotlib) for data visualization : network data, treemaps, charts

**User Experience Research/ Marketing Consultant**

**SolidWorks Corporation Consultant** (CAD software)

April 2007 —September 2010

Concord, MA

- ◆ Responsible for creation and tracking of user experience and satisfaction metrics for development organization, including executive dashboard design for the CTO
- ◆ Key data mining role on software quality projects: Identified patterns in software usage logs that suggest “crash” causes; design of data tools and dashboard to support stability research
- ◆ Established metrics and user feedback mechanisms on product installation, to guide design efforts and track improvements
- ◆ Refined corporate personas based on large quantitative and qualitative data collection in Europe that I managed personally; used cluster analysis and other data mining techniques for numeric data

**Platform Product Design Manager**

**Autodesk, Inc.** (architectural design software)

November 2005—January 2007

Waltham, MA

- ◆ Conducted quantitative research projects for product design and product management: Surveys, log analysis of product usage to inform design process and management (data mining done by me)
- ◆ Created processes, work deliverables, mentored to support new user-centered design focus for the division
- ◆ Hired for interaction design, visual design, contract work, and usability staff; managed external usability & design

**Principal Usability Specialist**

**The Mathworks, Inc.** (bio data tools)

May 2004—Oct 2005

Natick, MA

- ◆ Usability/design work for bioinformatics products, information visualization efforts for multi-dimensional data, and simulation products (joint patent pending from one project)

**Adobe Systems, Inc.** (Creative Suite products)

**Senior Interaction Designer**

Dec 2002—April 2004

Seattle, WA

- ◆ Interaction design spec authorship; mockups of workflow and dialog design for Creative Suite applications
- ◆ Cross-product design consistency for top priority complex customer problems: Color management, PDF workflow; these were “bullets on the box” for the CS2 release

**Director of Research and Development**

**Axance.com** (web usability)

March 2001—July 2002

Paris, France

- ◆ Expanded company expertise in user-centered design methodology and usability during product lifecycle

- ◆ Review and development of consulting proposals, methodology, test plans, reports, presentations; oversight of complex consulting projects; authorship of commercial proposals for international (multi-country) and more complex client projects
- ◆ Individual contributor on English-language projects for international clients, such as Renault and OpenTV; also provided expert consulting on several projects for Instant Messaging on mobile devices

**Manager, UI and Usability Group / Individual Contributor**

**TiVo, Inc.** (consumer electronics)

April 2000—December 2000

Alviso, CA

- ◆ Hiring (usability, UI, graphic designer and prototypers), employee reviews and general management
- ◆ Established relationships between market research efforts and usability/survey research
- ◆ Responsible for UI specification authorship, usability testing, & mentoring of junior usability staff

**Senior User Interface Designer**

**Excite, Inc.** (web portal)

Fall 1998—April 2000

Redwood City, CA

- ◆ Completed full design of three complex web applications, **Excite Planner** (calendar, address book, to do list, and notepad), **Excite Inbox** (mail and voicemail), **Excite Invite** (invitation service). Planner received industry awards for design, including favorable review in Forrester Report, and Inbox traffic increased by 30-50% after this redesign. (*Note: Excite has since been sold and products have changed substantially.*)
- ◆ Conducted some of the first in-house usability testing as part of design process; prepared and taught 1-day training course on UI and user-centered design process to product management staff (73 people)

**Member of Technical Staff, HCI Research Group**

**AT&T Labs Research** (spinoff from Bell Labs)

1995—1998

Florham Park, NJ

- ◆ Development of Internet chat system interfaces and study of online community: chat, email lists, newsgroups

**Part-time Employee/Consultant**

**SRI International**

Spring 1991—1993

Menlo Park, CA

- ◆ Made modifications to English grammar and lexicon for a natural language system with perl scripts

## SAMPLE TECHNICAL SKILLS and TOOLS

### Data Analysis / Mining

- ◆ Python Data Science tools (SpaCy, gensim, matplotlib, scikit-learn, Pandas, Dask, Spark, Jupyter, Tensorflow, PyTorch, etc.)
- ◆ R (ggplot, tidytext, Shiny, etc) – out of date!
- ◆ Tableau / Google Data Studio

### Other Tools...

- ◆ AWS / GCP experience
- ◆ Tensorflow / PyTorch / Huggingface tools
- ◆ Excel ☺
- ◆ Linux CLI
- ◆ Databases: SQLs, Spark, Redis, Mongo, BigQuery, Elasticsearch, Neo4j...
- ◆ Flask/FastAPI for Python APIs
- ◆ Some React / Svelte etc

## PATENTS

“**Viewer for Multi-Dimensional Data From a Test Environment**,” Patent awarded 2012 for work at The Mathworks (legal docket MWS-252), co-inventors James Owen, Jenifer Tidwell, Rong Chen, 2005.

“**Methods and Apparatus for Establishing Configuration Settings Between Applications**,” Patent awarded 2010, from work at Adobe (legal docket ADO05-07(B095)), co-inventors Lars Borg, Manish Kulkarni, Ken Kameda, 2005.

## OTHER PROFESSIONAL ACTIVITIES

Invited Keynote at [Wordplay: When Language Meets Games](#) workshop at NAACL 2022 ([talk slides](#))

Invited Jury Member for [Malofiej 2020](#), the Infographics World Conference (Society for News Design). (2020)

Runner-up in generative text competition by [Dead Alive Magazine](#), book forthcoming from New Site Press in 2021. (Work based on my National Novel Generation Month entry, "[Directions in Venice](#).")

Member of Advisory Panel for the "Visualizing Digital Scholarship in Libraries and Learning Spaces" grant funded by the Andrew W. Mellon Foundation at North Carolina State University (2017).

[Visiting Knight Foundation Chair](#) and teacher at University of Miami, Florida, 2015-2016.

Creator and administrator of [data-vis-jobs mailing list](#) for jobs in data visualization. (9 years!)

Former co-chair of well-regarded data visualization conference [OpenVis Conference](#) (5 years, conference now retired).

## WRITING EXAMPLES

Newsletter on creative AI/folklore/narrative at [arnicas.substack.com](#).

"[Plotto: Generating Offensive Stories Since 1928](#)." A data exploration of a book for narrative generation.

"[OpenVis Conference: Data Visualization on the Open Web](#)." An article about OpenVis Conference 2018.

"[COCO's Memory Palace](#)," a writeup of my artistic exploration of machine learning recognition of image content for poem generation at KIKK Festival in 2017.

"[Data Visualization 'vs' UI and Data Science](#)." An article responding to an online debate in 2017.

"[Teaching a Semester of D3.js](#)" A post about teaching D3.js to journalism students in Miami in 2016, which was widely read. (I think this is the one that ended up on Hacker News briefly.)

"[Visualizing Word Embeddings in Pride and Prejudice](#)," a post in 2014 about an "art" project that used word2vec on Jane Austen and had accompanying code/demo.

"[Data Characters in Search of an Author](#)," a 2014 post about data visualization insights.

"[Data Vis Consulting: Advice for Newbies](#)," a 2013 post that I think holds up and has input from colleagues.

## SAMPLE PUBLICATIONS / PUBLIC SPEAKING

- 2022 **NormConf 2022:** [NLP Tips & Tricks](#) talk on dim reduction, entity linking, classification with spaCy...
- 2022 **Invited Keynote,** [Wordplay: When Language Meets Games](#) workshop at NAACL 2022 ([talk slides](#) with video link)
- 2019 **PyData London Keynote Speaker** (London, July): "Getting Your Data Joie-de-Vivre Back." ([slides](#), [video](#))
- 2019 **EuroPython 2019 Keynote** (Basel, Switzerland, July) "Getting Your Data Joie-de-Vivre Back." ([slides](#))
- 2019 **Invited Keynote Panelist** on Data Visualization, **EMAE 2019 (Economics, Governance and Management of AI, Robots and Digital Transformations)** (Brighton, June) ([slides](#))
- 2018 **Keynote, PyData Warsaw** (November): "Tldr Summarization." ([slides](#))
- 2018 **Nantes "Data Week" Ecole de Design Invited Speaker** ([slides](#), [video](#))
- 2018 **Co-Chair and Organizer of OpenVis Conf 2018 in Paris** ([site](#))
- 2017 **KIKK Arts Festival: Invited Speaker on AI & Art** (Namur, Belgium) (written up [here](#))
- 2016 **PyData Paris & CNRS PyData Events in Rennes: Data Mining Job Offers** (Paris, Rennes, France)
- 2016 **Invited Speaker at DATO Data Science Conference** (San Francisco, July)
- 2016 **Joint Statistical Meetings (JSM) Invited Panelist** (on data visualization, July)
- 2016 **World Info Architecture Day Conference** (Miami, February, on text visualization)
- 2016 **Workshop on Text Analysis & Invited Speaker at Eyeo** (June, Minneapolis) ([video](#))
- 2015 **Carnegie Mellon U,** giving a visiting lecture on NLP and visualization for Golan Levin
- 2015 **NEASIST Conference** (Boston), talk on Interactive Data Visualization design

- 2014 **Magick Codes** (NYC unconference): An invited speaker on “Mining Dwarf Fortress for Stories”
- 2014 **The Graphical Web 2014 & The Bodleian (Oxford)**: Visualizing Topic Networks workshop
- 2014 **Reviewer for Pearson** for draft of **Visual Storytelling with D3** by Ritchie King
- 2013 **Harvard Librarians Data Science course**: taught 5 classes on Excel and visualization
- 2013 **Data Science Weeks**, Boston: talks on Interactive Vis, and Network Vis (in 2014)
- 2013 **OpenVis Conf** in Boston, talk “Bones of a Bestseller” (visualizing fiction)
- 2013 **Lean UX Conference**, NYC: Workshop on Web product analytics (intro to quant reasoning for usability)
- 2013 **PyData (Python Data) Silicon Valley 2013**: talk on “Nodebox” – for data visualization in python
- 2013 **Boston Predictive Analytics Machine Learning Tutorials**: Taught Ggplot2 and IPython notebook, organized statsmodels and sklearn tutorial sessions
- 2012 **Strata NYC talk**: “Beyond the Hairball” (network visualization techniques)
- 2012 **Boston Predictive Analytics**: “A Fast and Dirty Intro to NetworkX and D3.js”
- 2012 **Invited Speaker at Graduate UIUC HCI Talks Series**: “TV Fan Video Editing Network Evolution”
- 2010 **Boston Mini UPA conference**: “Mining Your Data: An Easy Intro to a Tough Topic”
- 2010 “I See Evidence of Research...What About Design?” **Article** in **<interactions>** sept + oct 2007
- 2008 “Designing ‘Up’ in the Software Industry.” An **essay** in **HCI Remixed**, collection edited by Tom Erickson and David McDonald, MIT Press.
- 2008 **Invited speaker at IBM Research**, Cambridge (spoke on TV fan video editing communities)
- 2006 **Closing Plenary Invited Speaker at Group ‘05 Conference** in Sanibel, Florida.
- 2001 **Conversation and Community: Chat in a Virtual World**, CSLI Publications. **My book** about Internet chat and virtual community based on my dissertation work.
- 1999 “The Dynamics of Mass Interaction,” by S. Whittaker, L. Terveen, W. Hill, L. Cherny. **Paper** presented at **CSCW 98**, Seattle, still widely cited online today.
- 1998 “Entertainment is a Human Factor: Game Design and CHI.” **Article** by L. Cherny, C. Clanton, E. Ostrom. **SIGCHI Bulletin**, October 1997. Summary of Workshop at CHI (Computer Human Interaction) 97.
- 1997 **Wired\_Women: Gender and New Realities in Cyberspace**, Seal Press. **Co-edited book** with Elizabeth Reba Weise. This was a popular textbook choice!

## OTHER WORK SAMPLES

- [ArXiv Text Generation Papers Search](#): a hand-rolled Jekyll site of crawled text generation papers on various topics of interest to me, updated once a week
- Monthly [newsletter on creative AI, narrative & games](#)
- [github.com/arnicas](https://github.com/arnicas) – not all of this is well structured or clean
- [Medium articles](#) on creative AI, data vis and analysis
- [ghostweather.slides.com/lynncherny](https://ghostweather.slides.com/lynncherny): Slide decks since 2016 on text analysis for non-programmers, exploring image collections with code, teaching AI, interactive data visualization, data sources “out there,” and more!
- [slideshare.net](https://slideshare.net) – older presentations and PPT format decks.
- “Visualizing Networks” talk for O’Reilly Strata 2012 is on their Slideshare site with 87K views.

## EDUCATION

- 1990–1995 **Ph.D.** in Linguistics (with courses in HCI, computer science, ethnography)  
**Stanford University, CA**  
Dissertation: *The MUD Register: Conversational Modes of Action in a Text-Based Virtual Reality*. (Ethnography of an Internet "MUD" text chat community, later published as book available on Amazon)
- 1989–1990 **M.Phil.** in Computer Speech and Language Processing  
**Cambridge University (Trinity College), UK**
- 1986–1989 **B.A.** in Linguistics (with courses in AI at Edinburgh University)  
**University of Maryland, College Park, MD**