Leveraging Deep Learning and Natural Language Processing for Publication Readiness

September 28, 2020

1100 GMT

Nishchay Shah
nishchay.shah@cactusglobal.com

CACTUS | LABS

A CACTUS Solution

Research and Writing

Language Compliance

Technical Compliance

CACTUS | LABS

A CACTUS Solution

Automated Language Editing OR Grammatical Error Correction (GEC)

Traditional, Rule Based

- Handwritten, linguistically influenced rules
- Hard to scale, difficult to maintain
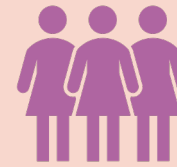- Limited contextual coverage
- High accuracy, very low coverage

CACTUS: Neural Network Powered

- State of the art neural networks
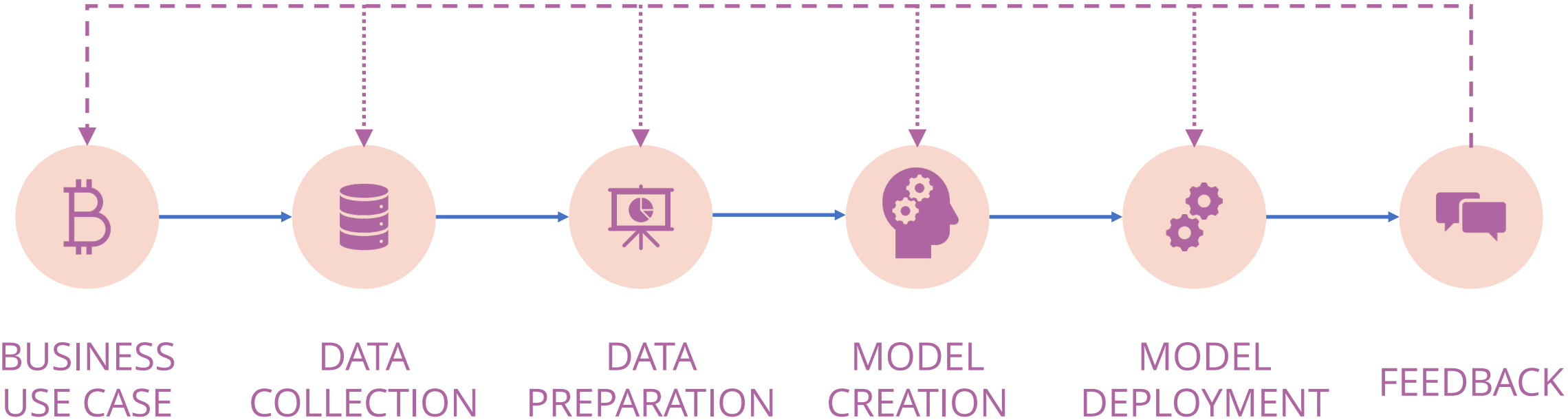- Trained on edits curated from over 100K academic authors; 300M+ sentences
- Long context memory, large semantic knowledge
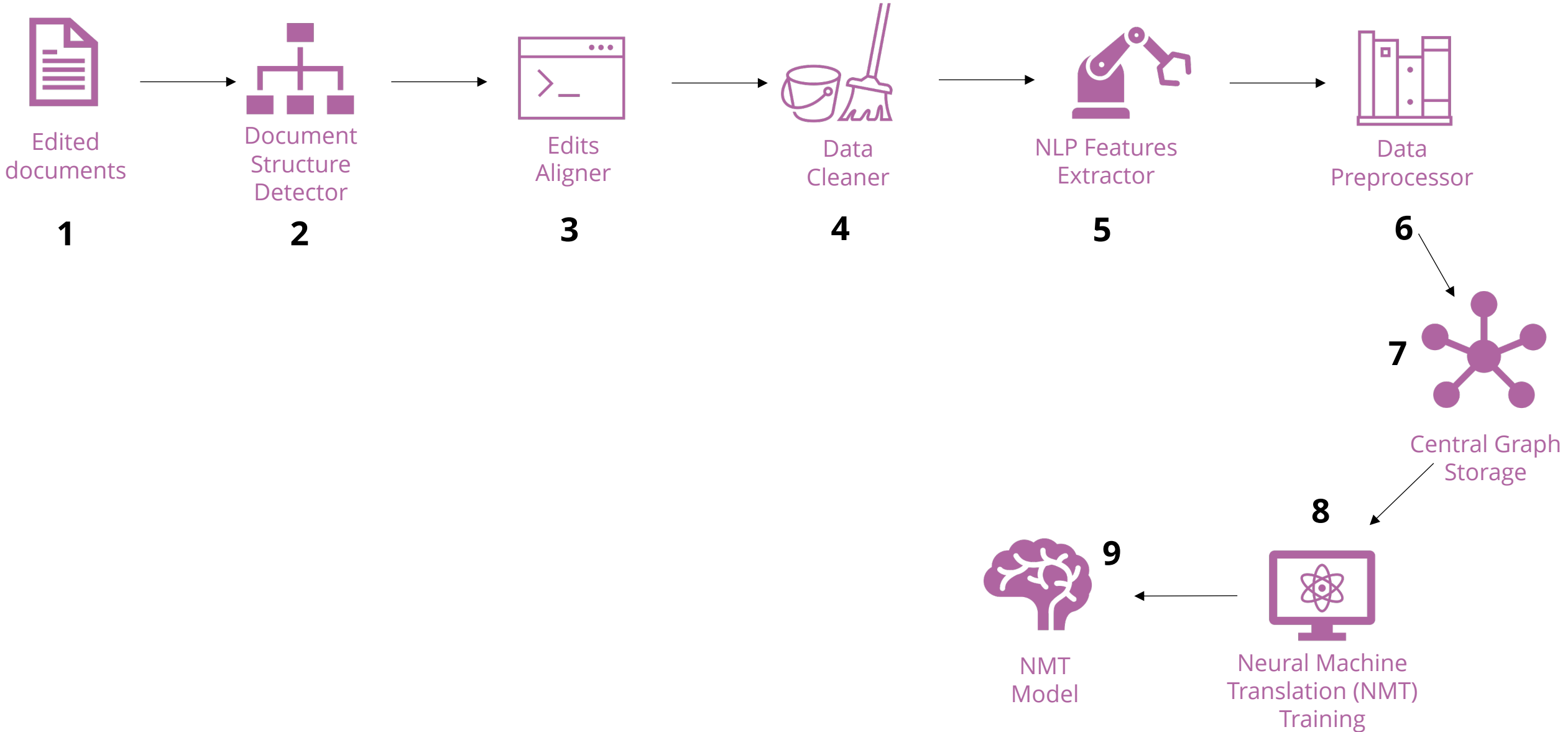- High accuracy with high coverage

CACTUS | LABS

A CACTUS Solution

BUSINESS USE CASE → DATA COLLECTION → DATA PREPARATION → MODEL CREATION → MODEL DEPLOYMENT → FEEDBACK
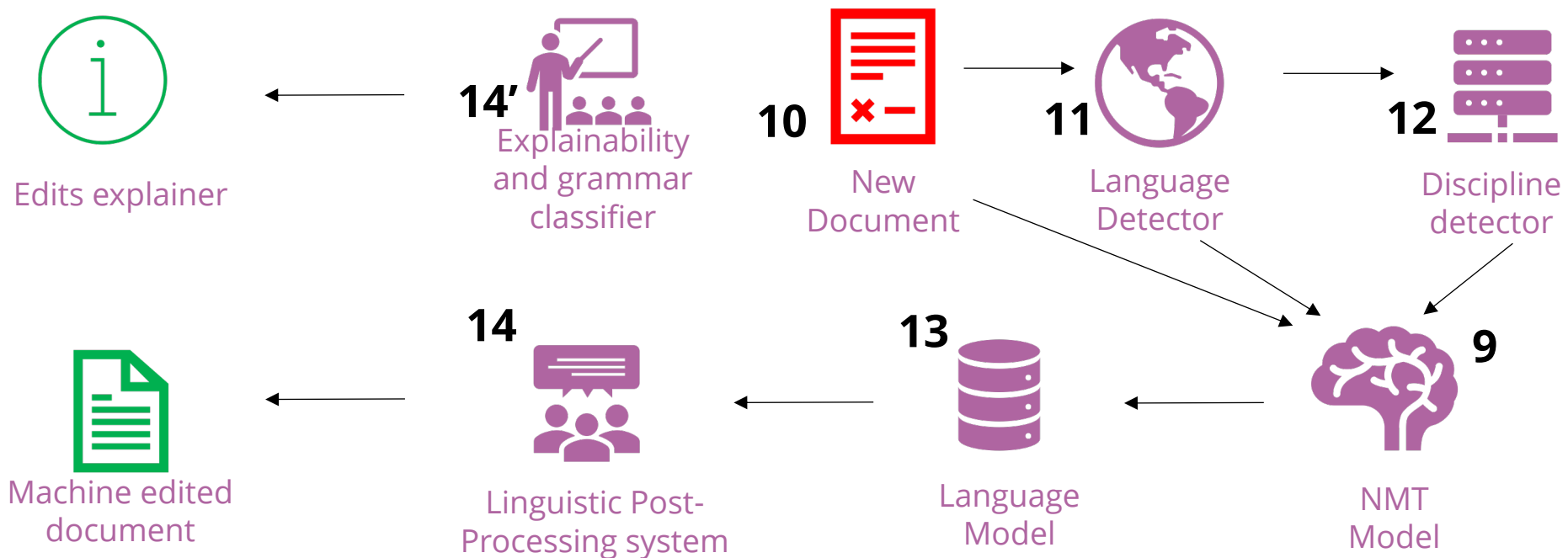
Neural Machine Editing Model In Production

ORIGINAL SENTENCE	MACHINE EDITED	HUMAN EDITED
There is no vaccine or drugs for ASFV available, and control method include preventive biosecurity and mass slaughter in the case of an outbreak.	There are~~is~~ no vaccines~~vaccine~~ or drugs available for ASFV~~for ASFV available~~, and control ~~method~~ methods include preventive biosecurity and mass slaughter in the case of an outbreak.	~~There is n~~No vaccines or drugs are available for ASFV ~~available~~, and the control method includes preventive biosecurity and mass slaughter in the case of an outbreak.
The efficacy of protection of two adenovirus vectored ASFV antigen cocktails was evaluated as an extension of previous work which performed an evaluation of safety, tolerability, and immunogenicity of these immunogens.	The efficacy of protection of two adenovirus-vectored~~adenovirus vectored~~ ASFV antigen cocktails was evaluated as an extension of a previous study~~work~~ that evaluated the~~which performed an evaluation of~~ safety, tolerability, and immunogenicity of these immunogens.	The ~~efficacy of~~ protection efficacy of two adenovirus vectored ASFV antigen cocktails was evaluated as an extension of a previous work in which ~~performed an evaluation of~~the safety, tolerability, and immunogenicity of these immunogens was evaluated.
As far as we know, this compilation is the first attempt to shed light on the shortcomings as more and more studies on molecular hybridization of the active pharmacophores of known antimalarials are being published.	To the best of our knowledge~~As far as we know~~, this ~~compilation~~ is the first attempt to shed light on the shortcomings of an increasing number of~~as more and more~~ studies on molecular hybridization of the active pharmacophores of known antimalarials are being published.	As far as we know, this ~~compilation~~ is the first ~~attempt~~ report that ~~to~~ sheds light on these shortcomings, even as ~~more and more~~a growing number of studies are reporting ~~on~~ the molecular hybridization of the active pharmacophores of known antimalarials ~~are being published~~.

CACTUS | LABS

A CACTUS Solution

1 Edited documents

2 Document Structure Detector

3 Edits Aligner

4 Data Cleaner

5 NLP Features Extractor

6 Data Preprocessor

7 Central Graph Storage

8 Neural Machine Translation (NMT) Training

9 NMT Model

1 Edited documents

2 Document Structure Detector

3 Edits Aligner

4 Data Cleaner

5 NLP Features Extractor

6 Data Preprocessor

7 Central Graph Storage

8 Neural Machine Translation (NMT) Training

9 NMT Model

10 New Document

11 Language Detector

12 Discipline detector

13 Language Model

14 Linguistic Post-Processing system

14' Explainability and grammar classifier

Edits explainer

Machine edited document
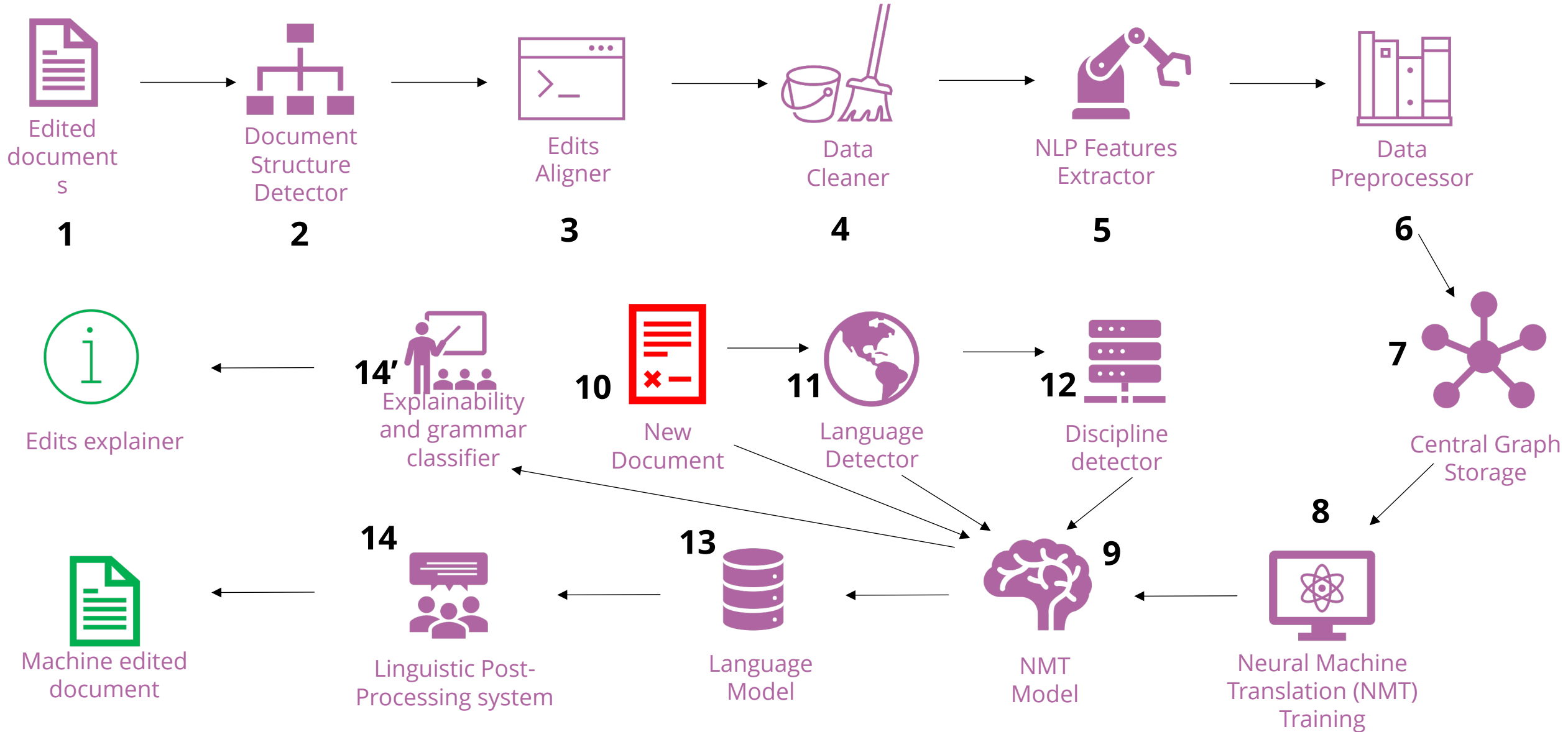
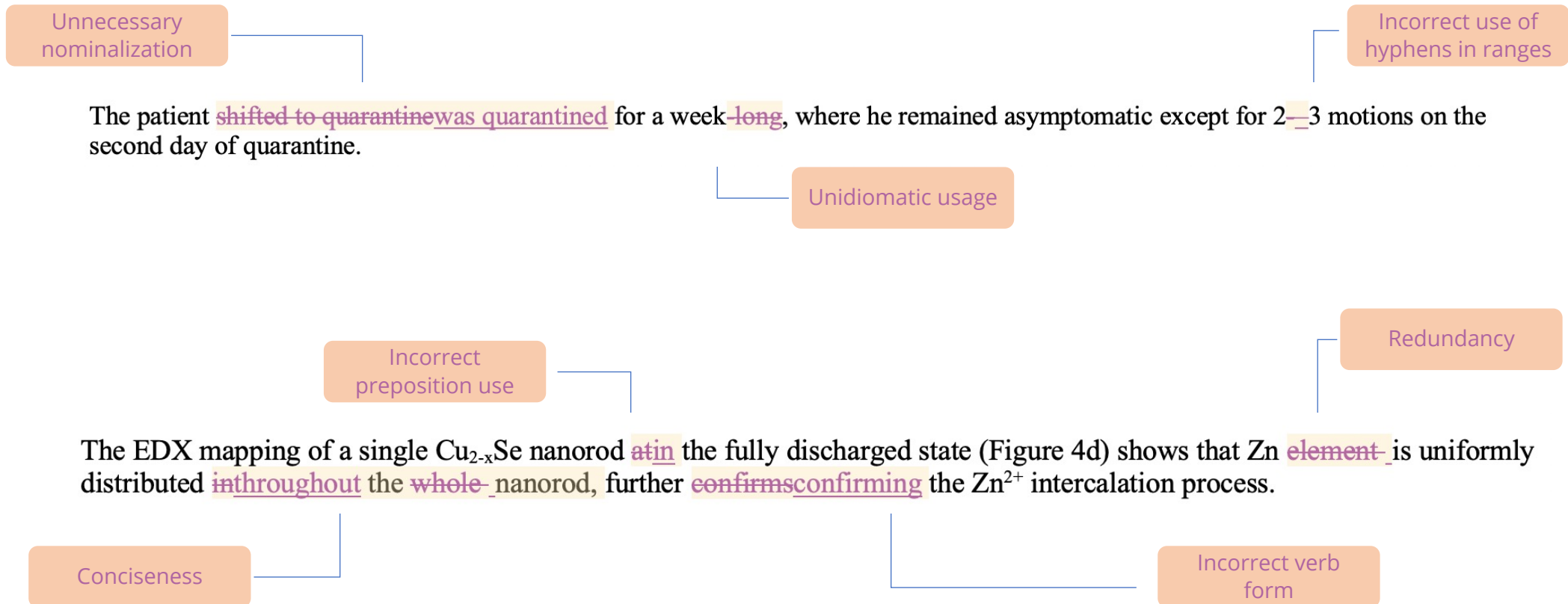The neural language polishing solution at CACTUS does more than a grammar check.

It provides context-sensitive suggestions for vocabulary, style, and academic tone.

Unnecessary nominalization

Incorrect use of hyphens in ranges

The patient ~~shifted to quarantine~~was quarantined for a week ~~long~~, where he remained asymptomatic except for 2~~–~~3 motions on the second day of quarantine.

Unidiomatic usage

Incorrect preposition use

Redundancy

The EDX mapping of a single $Cu_{2-x}Se$ nanorod ~~at~~in the fully discharged state (Figure 4d) shows that Zn ~~element~~ is uniformly distributed ~~in~~throughout the ~~whole~~ nanorod, further ~~confirms~~confirming the Zn^{2+} intercalation process.

Conciseness

Incorrect verb form

CACTUS | LABS

A CACTUS Solution

10

Contextual word choice

Redundancy

The linearity of fungicides was evaluated at ~~serious~~various ~~of concentration levels~~ concentrations (0.01, 0.05, 0.1, 0.2, 0.5, 1, 2, and 5 µg/L).

Informal word usage

Semicolon between independent clauses

Emerging evidence indicates that MVI is an independent factor for resection of HCC with poor survival[20-22]~~,~~; however, MVI is ~~hard~~difficult to detect before partial hepatectomy, even if ~~the improvement of~~ superior imaging procedures have been improved in recent years.

Rephrase for readability

Noun number

Adsorption technology has been studied as a promising ~~candidate~~alternative to ~~replace the~~ traditional C3 separation ~~unit~~units.

Wordiness

Wordiness

Incorrect preposition

When ~~we apply it~~applied to the skin, the paste reacts with the water ~~of~~on the skin surface, generating CO_2 gas.

Missing article

Awkward phrasing

Obviously, ~~with~~as the ~~increasing of user~~ number of users increases, the total utility of the users ~~is becoming higher~~increases.

Unnecessary Nominalization

In the present study, we ~~conducted an evaluation of~~evaluated the protective efficacy of both cocktails by intranasal challenge of pigs with ASFV-Georgia 2007/1.

Missing article

CACTUS | LABS

A CACTUS Solution

2.5 Billion+ words edited in last 12 months

Authorship

Checks for corresponding author email

Commercial Declaration

Declarations of conflicts of interest (financial/non-financial)

Ethical Declaration

The author should declare whether written informed consent was obtained from human subjects

Reference Check

Proper referencing of other works used to build on the study

Tables & Figures

Every visual element included in the manuscript should be mentioned in the text

Ethical Use of Images

Privacy should be respected in Images of human participants

Inclusive Language

Improper use of language that can marginalise, demean, and exclude social groups such as women, etc

Structure

Follow a standard structure in which information is presented

Self-citation Checks

Including a large number of self-citations in a manuscript is discouraged

Data Access Statements

Indicate general data reporting compliance

Citations in Abstract

Avoid citing sources in the abstract of your manuscript

Copyrighted Methods

Copyrighted methods mentioned in the manuscript are highlighted

Journal Recommendations

Top recommendations that are a match between journal and the manuscript top concepts

Journal Recommendations
Top recommendations that are a match between journal and the manuscript top concepts

85% Match | EXCELLENT
Phi Delta Kappan

Top concepts matching papers in the journal and your manuscript

Accountability

Show papers matching your manuscript

Report incorrect recommendation

85% Match | EXCELLENT
Accounting, Auditing & Accountability Journal

Top concepts matching papers in the journal and your manuscript

Accountability

Show papers matching your manuscript

Report incorrect recommendation

Assistive AI

The early days

High investment

High uncertainty

Many failures and do-overs

Augmented AI

The scale up

Pushing through the resistance

Large iterations

Autonomous AI

Hyper Automation

Big bucks

Significant displacement

Model and success replication

CACTUS|LABS

A CACTUS Solution

nishchay.shah@cactusglobal.com

CACTUS | LABS

A CACTUS Solution