

Application of AI in Drug Development

JULIA CHEN

STATISTICAL PROGRAMMING MANAGER, HUTCHISON MEDIPHARMA

Outlines

- A brief – AI Evolution
- A Use Case – Facilitate and Advance Submission Package
 - TLF Review
 - Documents Checking
- A Practical Example – Image Analysis
- Opportunities and Challenges

AI Evolution



- | | | | | | | |
|--|--|--|---|---|--|---|
| <ul style="list-style-type: none">• 1952–1956• Cybernetics & early neural networks• Turing's test• Dartmouth Conference | <ul style="list-style-type: none">• 1956–1974• Natural Language• The Optimism• The Money: MIT received 2.2 million grant• Robotics | <ul style="list-style-type: none">• 1974–1980• Limited computer power and space• Common sense and knowledge• The End of Money• Criticism from across campus• Minsk/Shank: Object-oriented programming | <ul style="list-style-type: none">• 1980–1987• The rise of expert system• The knowledge revolution• The Money returns• Robotics | <ul style="list-style-type: none">• 1987–1993• Market of AI hardware collapse• cut funding to AI "deeply and brutally." | <ul style="list-style-type: none">• 1993–2011• Milestones and Moore's law• Deep Blue in Chess• Victory of the NEATS• Algorithm developed | <ul style="list-style-type: none">• 2011–Now• Deep Learning• Big data• Full AI |
|--|--|--|---|---|--|---|

AI in drug development

AI for **drug target identification and validation**

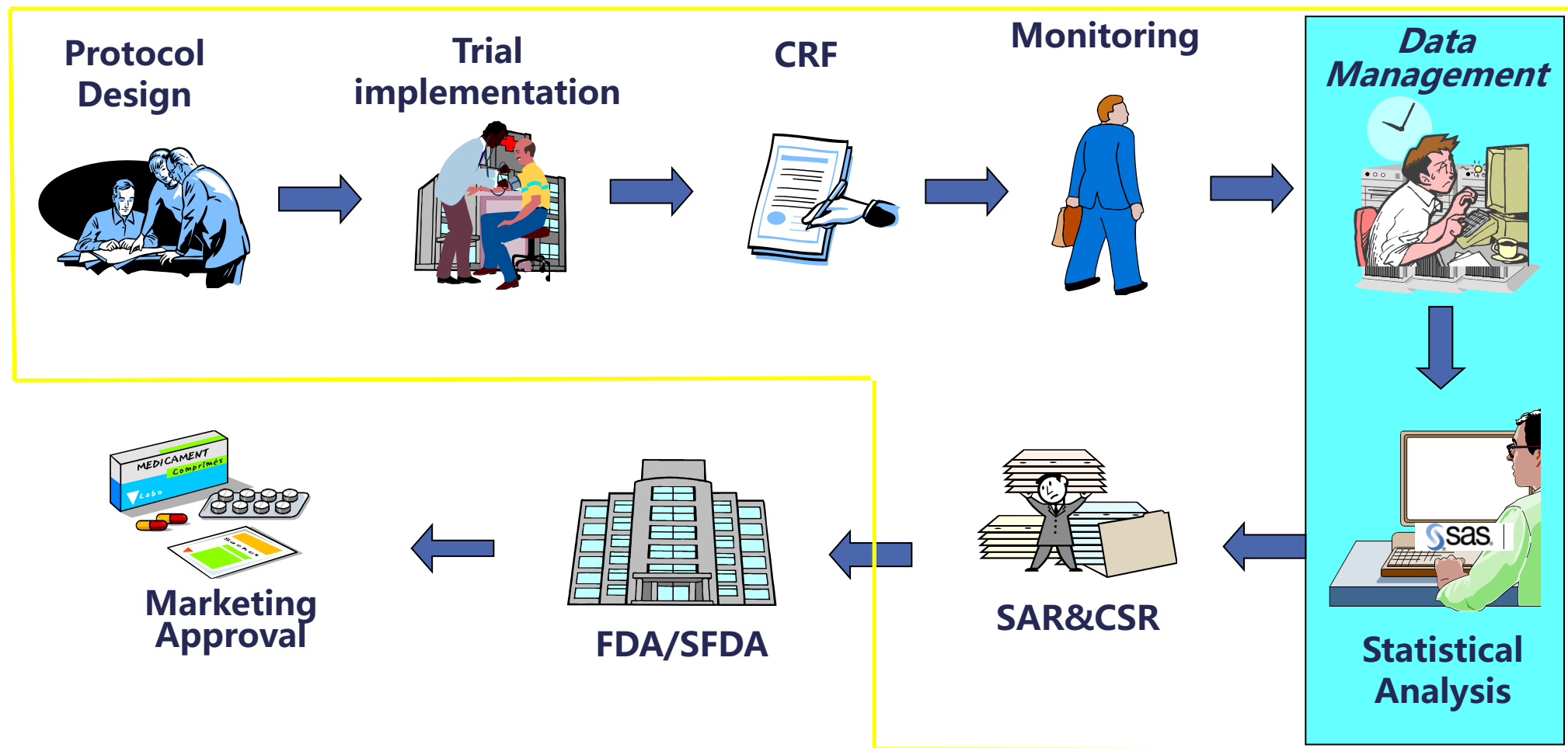
AI for **biomarkers development**

AI for **dealing with biomedical, clinical and patient data**

AI for **drug repurposing programs**

AI for **analyzing research literature, publications, and patents**

Clinical Trial Overview





REPEATED, TRIVIAL, TIME-CONSUMING

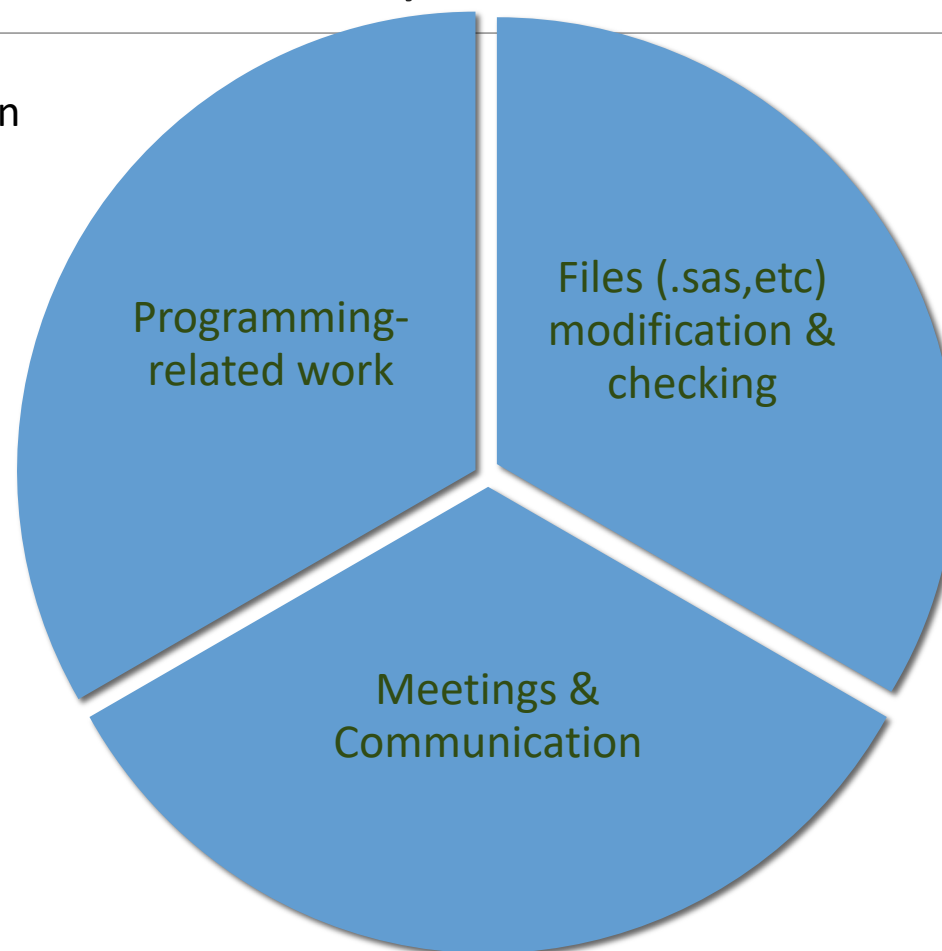
SOLUTION: THE IDEA OF NLP

Natural Language Processing

- **Natural language processing (NLP)** is a subfield of [computer science](#), [information engineering](#), and [artificial intelligence](#) concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze [natural language](#) data.
 - Morphological segmentation
 - Word segmentation
 - Machine translation
 - Optical character recognition
 - Speech recognition

How to Improve Efficacy

- Dataset/Spec creation
- TLF generation
- Macro testing
- Issue checking
- Tools:
Spotfire/JMP,etc



- Create folders
- Move files from study to study
- Modify program name/contents
- Check files from CRO/Internal
- Check validation passed
- Check headers/Footnotes
- Check E-Submission package

Use Case – Check Suspected Misspelling in TLFs

- **Common issue in TLF:** Typos, unrecognized character, upper/lowercase, extra space, etc
- **Demo: Check the string:** - 'E/N: Event number/ Subject number.'

```
%macro Tlf_Rev(string = ,
               %SYSEXEC \home\zhuoc\python\tlf_rev.py %input_param ([string: &string.])
%mend;

*Call the macro;
%Tlf_Rev(string = C: \Users\zhuoc\Desktop\Demo\Compound123\Final\Dev\Documents\check_word.txt);
```

- **Get below results:**

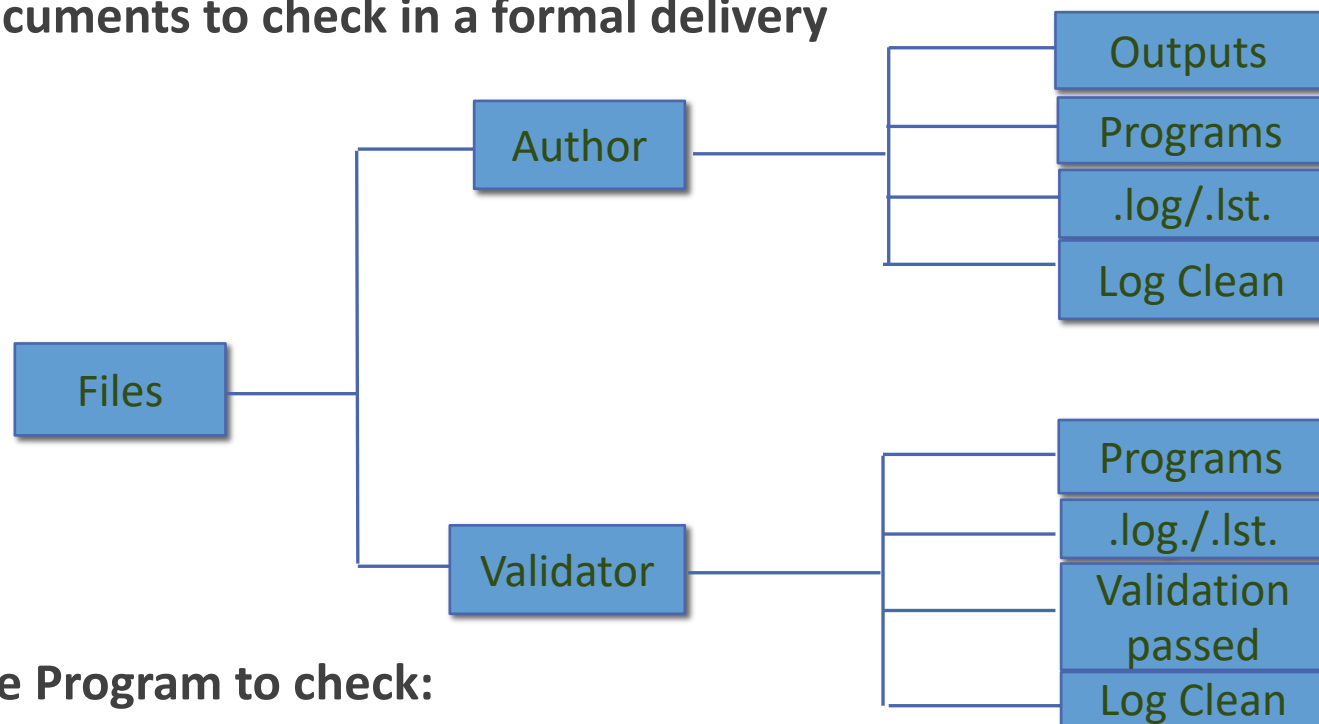
```
/Demo/Compound123/Final/Dev/Outputs/Author/Tables/t_14.2.5.4.rtf
"SOURCE: Listing 16.2.19.2
Page 1 of 4 /N: Event number/ subject number.
EGF: Vascular Endothelial Growth Factor. EGFR: Epidermal growth factor receptor. edian overall survival time and 95% CI are estimated by Kaplan-Meier method,
Hazard Ratio and 95% CI are calculated rom unstratified Cox proportional hazard model. ensoring rule is the same as which was used for PFS primary analysis.
] First diagnosis is ileocencal malignant tumor for 30091010, thus no applicable primary site was selected. :\Hutchison\Fruquintinib\EXA29677\Biostatistics\P
roduction\Tables\Final\T_14.2-5.4.sas 28APR2017 10:29" has error str
in term of spelling: "E/N: Event number/ Subject number."
/Demo/Compound123/Final/Dev/Outputs/Author/Tables/t_14.2.1.7.rtf
"SOURCE: Listing 16.2.15
Page 1 of 1 /N: Event number/ Subject number %
Percentage of subjects in each category relative to the total number of subjects in the relevant analysis set, with assessments available. A subject can ha
ve more than one category of tumor type and assessment method. :\Hutchison\Fruquintinib\EXA29677\Biostatistics\Production\Tables\Final\T_14.2-1.sas 28APR2017
10:28" has error str
in term of spelling: "E/N: Event number/ Subject number."
```


Other usage: Recognize free texting in CRF, SAE Report, medical case, etc.

```
if __name__ == '__main__':
    main()
```

Use Case - Documents checking

Documents to check in a formal delivery



Use Program to check:

```

%macro Doc_Chk(input = ,
               output = ,
               %SYSEXEC \home\zhuoc\python\doc_chk.py %input_param ([input: &input.] [output: &output.])
%mend;

*Call the macro;
%Doc_Chk(input = C: \Users\zhuoc\Desktop\Demo\Compound123\Final\Dev\Documents\ADaM_Chk.xlsx,
         output = C: \Users\zhuoc\Desktop\Demo\Compound123\Final\Dev\Documents\Chk_Rslt.xlsx);
  
```

Use Case - Documents checking



Input Excel

ADaM Datasets
ADSL
ADDS
ADAE
ADMH
ADCM
ADEG
ADVS
ADEX
ADEXS
ADTR
ADRS
ADTTE



Output Excel

ADaM Datasets	Output	Author Prg	Author Log/Lst	Validator Prg	Validator Log/Lst	Validation passed?	Log clean?
ADSL	Y	Y	Y	Y	Y	Y	Y
ADDS	Y	Y	Y	Y	Y	Y	Y
ADAE	Y	Y	Y	Y	Y	Y	Y
ADMH	Y	Y	Y	Y	Y	Y	Y
ADCM	Y	Y	Y	Y	Y	Y	Y
ADEG	Y	Y	Y	Y	Y	Y	Y
ADVS	Y	Y	Y	Y	Y	Y	Y
ADEX	Y	Y	Y	Y	Y	Y	Y
ADEXS	Y	Y	Y	Y	Y	Y	Y
ADTR	Y	Y	Y	Y	Y	Y	Y
ADRS	Y	Y	Y	Y	Y	N	N
ADTTE	Y	Y	Y	Y	N	N	N

Use Case - Documents checking

Function: Check the existence of required documents

Language: Python

Working environment: Windows or Unix/Linux

Logic:

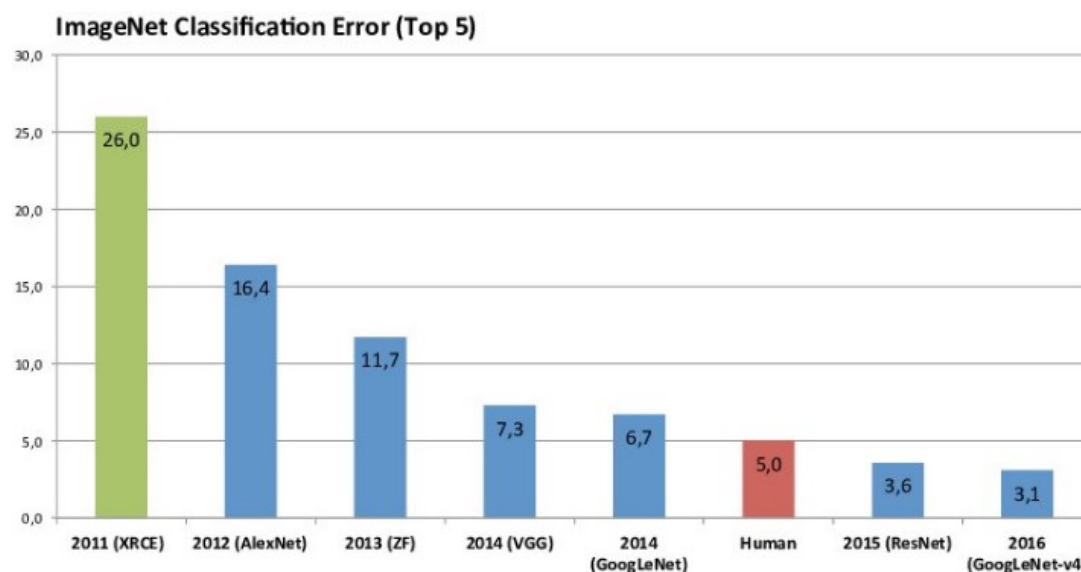
- 1: Create the input excel with desired ADaM names
- 2: Check if the corresponding files exists, if exists then "Y", else "N"
- 3: Check if the searching words "No unequal values were found. All values compared are exactly equal." in validation log, and "No searched messages found" in lst, then label "Y/N"
- 4: Input the read-in excel path/name and result path/name in Python ,run the program.

Advantages: 1) Run in batch

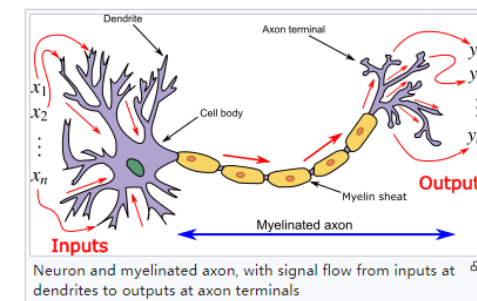
2) More complicated with experience accumulate

A Practical Example

- Computer Vision: Deals with how computers can be made to gain high-level understanding from digital images or videos.
- Deep Learning



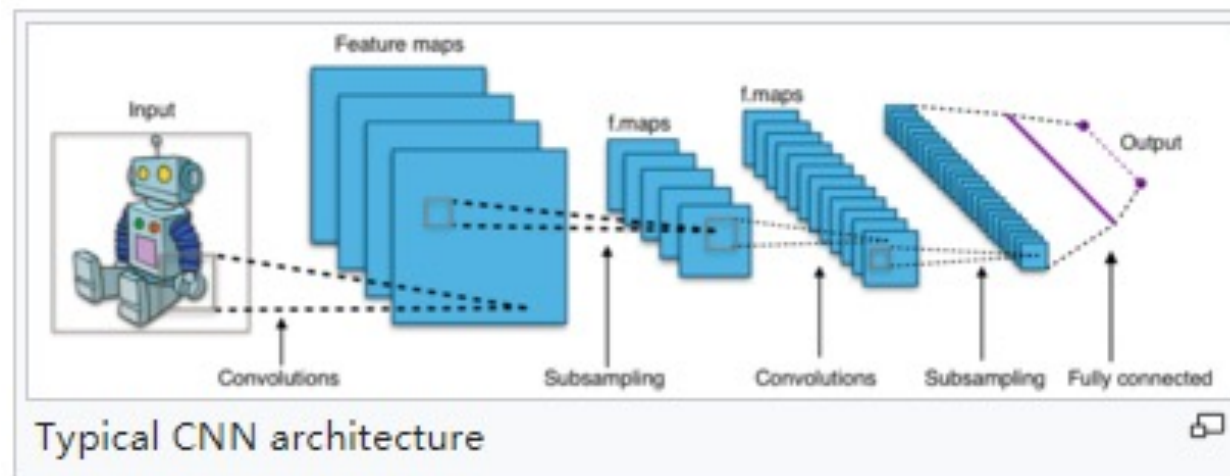
- CNN: Convolutional Neural Network



A Practical Example – Image Analysis

Key Concepts about CNN:

- Loss Functions and Optimization
- Data augmentation
- Filter and other parameters
- Convolutional layers
- Detection and Segmentation



- In recent Kaggle campaign , among 888 scans of the publicly available LIDCIDRI dataset, the best method reaches high detection sensitivities of 85.4% and 90.1% at 1 and 4 false positives per scan, respectively. The result is better than human judgment.

Opportunities and Challenges

Opportunities:

- Pharmaceutical industry is booming with the high- developed technologies and investment.
- More job opportunities created
- Different career development

Challenges:

- Learning curve
- Real Operation

Further Information



Speaker E-mail Address: zhuoc@hmplglobal.com

 *PhUSE China*
 *phuse.eu*
 *[@PhUSETwitta](https://twitter.com/PhUSETwitta)*
 *[/PhUSE](https://www.facebook.com/PhUSE)*
 *phusewiki.org*

