**Narrative Extraction Plan**

There are a few aspects of the letters that we are interested in extracting using GPT4 and other tools. For each piece of information, we will want to tag the index/indices of the sentence(s) that provide that information.

*Addressee & the Opportunity to Write*

To whom is the letter addressed? We want to select all of the following that apply:

* Warren Buffett
* Doris Buffett
* Sunshine Lady Foundation
* Letters Foundation
* No specific addressee (e.g. "To Whom It May Concern")
* Other (5) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(We might also consider extracting the address itself and sorting it into one of several addresses to which the Buffetts tended to receive letters, but this seems of secondary importance.)

How did the author learn about the opportunity to write?

* [ could probably come up with a typology of media channels based on what we’ve read / on Mane’s Buffett press timeline]

*Actors and Relations*

We hope to pull out the actors described in the letter, their pertinent attributes, and their relationships.

One way to do this may be to assign numbers (or letters) to actors in the text. The author is 1, and subsequent actors are 2, 3, 4 and so on. For each actor, we will want to capture:

* Gender
* Age, if mentioned
* Roles / relations to other actors mentioned, from among the following categories:
  + Kin relations
    - Spouse/partner
    - Ex-spouse/partner
    - Parent
    - Child
    - Sibling
    - Grandparent
    - Grandchild
    - Aunt/Uncle
    - Niece/Nephew
    - Cousin
    - Widow(er)
  + Other relations:
    - Boss / manager
    - Friend
    - Neighbor
    - Colleague /work associate
    - Student
    - Teacher

Then, there are other roles and identities that people use to describe themselves, without discussion of specific interpersonal ties.

* Christian
* Local volunteer
* Longtime resident
* Veteran
* Citizen
* Immigrant
* American
* Entrepreneur, business owner
* Widow(er)
* Divorcee
* Orphan
* Single parent
* Retiree
* Homemaker
* Student

We may also want to capture occupation (could sort into a few different categories) and/or income if mentioned.

We will tag all of the information above with the indices of sentences in which that information appears.

Ideally for each actor, we would also list all sentence indices (and positions within those sentences) corresponding to mention of that actor, whether or not the sentence provides information included above. Then we might later take advantage of this data and mine these sentences for other information about character attributes, relationships between characters, and so on.

***Events, Cognitions, Macro-events and States/Conditions***

For each event or state/condition mentioned in the letter that corresponds to one of the following categories (from the Qualtrics form) and associated codes, we will ask GPT4 to list a) the event or state/condition mentioned, b) the code for that event or state/condition, and c) the index of the sentence(s) that describe that state and/or condition.

There may be a better coding scheme, but here’s the arbitrary one I assigned to test GPT4:

Housing-related events

- (Threat of) eviction or foreclosure: 2\_1\_1

- Housing market decline or 'going under': 2\_1\_2

- Housing damage: 2\_1\_3

- Other housing event: 2\_1\_4

Employment-related events

- lost job or laid off: 2\_2\_1

- fired: 2\_2\_2

- work hours cut or lost/reduced business: 2\_2\_3

- recession: 2\_2\_4

- other employment event: 2\_2\_5

Family or household-related events

- divorce: 2\_3\_1

- separation: 2\_3\_2

- new dependents: 2\_3\_3

- lost wage earner: 2\_3\_4

- abuse: 2\_3\_5

- addiction / substance abuse: 2\_3\_6

- death: 2\_3\_7

- other family or household-related event: 2\_3\_8

Medical events

- traumatic injury: 2\_4\_1

- sudden illness: 2\_4\_2

- cancer: 2\_4\_3

- diagnosis: 2\_4\_4

- surgery or other medical treatment requiring recovery: 2\_4\_5

- other medical event: 2\_4\_6

Transportation events

- vehicle breakdown: 2\_5\_1

- vehicle repossessed: 2\_5\_2

- vehicle impounded: 2\_5\_3

- vehicle stolen: 2\_5\_4

- other transportation-related event: 2\_5\_5

incarceration 2\_6

natural disaster 2\_7

other event 2\_8

States and conditions

- (history of) poverty: 3\_1

- unemployment: 3\_2

- unable to work: 3\_3

- no health insurance: 3\_4

- credit card debt accumulation: 3\_5\_1

- mortgage debt accumulation: 3\_5\_2

- school loan debt accumulation: 3\_5\_3

- medical debt accumulation: 3\_5\_4

- general debt accumulation: 3\_5\_5

- other debt accumulation: 3\_5\_6

- (history of) domestic violence: 3\_6

- chronic illness and/or recurring medical issues: 3\_7

- disability: 3\_8

- mental or psychological issues: 3\_9

- addiction or substance abuse: 3\_10

- other state or condition: 3\_11

Note: Above we have a typology that applies specifically to adverse events. I wonder whether we might want/need to list out types of “cognitive events” like those included in Becoming a Nazi.

*Causal Sequences*

Then, for each pair of events A and B, choose one of the following statements: A causes B, B causes A, or, if neither, move to the next pair. (Note: learning about the charity opportunity and writing the letter are not events.)

*The Need(s)*

Then, for each present, unmet need mentioned in the letter that corresponds to one of the following categories and associated codes, list a) the need, b) the code corresponding to that need, and c) the index of the sentence(s) that describe that need. Here are the categories (again from Qualtrics) and their codes:

- food: 5\_1

- clothing: 5\_2

- to pay bills (general): 5\_3

housing and housing-related needs

- housing (general): 5\_4\_1

- to pay rent: 5\_4\_2

- to pay mortgage: 5\_4\_3

- to relocate: 5\_4\_4

- utilities: 5\_4\_5

- home repairs: 5\_4\_6

- home/property renovation or expansion: 5\_4\_7

- accessible housing: 5\_4\_8

- to purchase new home or land: 5\_4\_9

- other housing need(s): 5\_4\_10

medical need

- dental care: 5\_5\_1

- surgery: 5\_5\_2

- medication: 5\_5\_3

- physical therapy: 5\_5\_4

- psychotherapy and/or psychiatry: 5\_5\_5

- general medical care: 5\_5\_6

- other medical need: 5\_5\_7

debt relief

- general debt relief: 5\_6\_1

- credit card debt relief: 5\_6\_2

- mortgage debt relief: 5\_6\_3

- school loan debt relief: 5\_6\_4

- medical debt relief: 5\_6\_5

- other debt relief: 5\_6\_6

- a job: 5\_7

- (to pay for) education / training: 5\_8

- school / learning supplies: 5\_9

- to go on a trip (vacation, to visit family): 5\_10

- to better care for / help family: 5\_11

- to pay for funeral expenses: 5\_12

- to start a business: 5\_13

- to start a civic organization: 5\_14

transportation and transportation-related needs

- vehicle repair: 5\_15\_1

- accessible vehicle: 5\_15\_2

- other new (or used) vehicle: 5\_15\_3

- other transportation needs: 5\_15\_4

- other needs: 5\_16

If the author notes a specific monetary cost to meet the need, then list it after the code.

*The Ask(s)*

3) Finally, for each ask that the author makes in the letter, list a) the type of ask, b) the code corresponding to that ask, and c) the index of the sentence(s) that describe that ask. Here are the Qualtrics categories and their codes:

- Specific monetary amount 7\_1

- Loan 7\_2

- help with specific needs mentioned 7\_3

- help, unspecified / general 7\_4

- advice 7\_5

- other ask 7\_6

If the ask is for help with one or more of the specific needs listed earlier, then please list the needs the author asks for help with. If the ask is for a specific monetary amount (whether a donation or a loan) then we also want that $ amount.

Ideally we would also list on whose behalf each ask is made.

*The Future*

This is not in our form right now, but I’m not sure whether our current plan of information extraction captures information that authors provide about bridges to the future, which are often distinct from the ask itself. That is, what will the requested help do for them. For further thought / discussion.

*Other sources of support*

I have yet to modify this Qualtrics question for GPT4, but we have a question about whether the writer explicitly mentions the possibility of other support for the present need (whether or not that support has been obtained), and if so, what kind(s) of support? The options are:

* Family support
* Government support
* Community support
* Bank loan or loan renegotiation
* Other (fill in) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

We then ask whether the intended beneficiary has received any of this support.

*Attachments and/or references*

Finally, we ask whether the author provides references (that is, contacts who might corroborate her story), and whether she provides attachments.