|  |  |  |  |
| --- | --- | --- | --- |
| File Name | ToM Dataset | Question Type | Testing Result |
| FANToM\_FactQ.csv | FANToM | FactQ in FANToM | A scalar value between 0.0 and 1.0, indicating the token level F1 score based on the response and the correct answer. Higher is better. |
| FANToM\_BeliefQ\_Choice\_First\_Order.csv | FANToM | First-Order BeliefQ [Choice] in FANToM | A binary value, i.e., True or False, indicating whether the choice is correct or not. |
| FANToM\_BeliefQ\_Choice\_Second\_Order\_Cyclic.csv | FANToM | Second-Order (Cyclic) BeliefQ [Choice] in FANToM | A binary value, i.e., True or False, indicating whether the choice is correct or not |
| FANToM\_BeliefQ\_Choice\_Second\_Order\_Acyclic.csv | FANToM | Second-Order (Acyclic) BeliefQ [Choice] in FANToM | A binary value, i.e., True or False, indicating whether the choice is correct or not |
| FANToM\_BeliefQ\_Dist\_First\_Order.csv | FANToM | First-Order BeliefQ [Dist.] in FANToM | A binary value, i.e., True or False, indicating whether the response is more similar to the correct answer or the wrong answer. |
| FANToM\_BeliefQ\_Dist\_First\_Order\_TokenF1.csv | FANToM | First-Order BeliefQ [TokenF1] in FANToM | A scalar value between 0.0 and 1.0, indicating the token level F1 score based on the response and the correct answer. Higher is better. |
| FANToM\_BeliefQ\_Dist\_Second\_Order\_Cyclic.csv | FANToM | Second-Order (Cyclic) BeliefQ [Dist.] in FANToM | A binary value, i.e., True or False, indicating whether the response is more similar to the correct answer or the wrong answer. |
| FANToM\_BeliefQ\_Dist\_Second\_Order\_Cyclic\_TokenF1.csv | FANToM | Second-Order (Cyclic) BeliefQ [TokenF1] in FANToM | A scalar value between 0.0 and 1.0, indicating the token level F1 score based on the response and the correct answer. Higher is better. |
| FANToM\_BeliefQ\_Dist\_Second\_Order\_Acyclic.csv | FANToM | Second-Order (Acyclic) BeliefQ [Dist.] in FANToM | A binary value, i.e., True or False, indicating whether the response is more similar to the correct answer or the wrong answer. |
| FANToM\_BeliefQ\_Dist\_Second\_Order\_Acyclic\_TokenF1.csv | FANToM | Second-Order (Acyclic) BeliefQ [TokenF1] in FANToM | A scalar value between 0.0 and 1.0, indicating the token level F1 score based on the response and the correct answer. Higher is better. |
| FANToM\_AnswerabilityQ\_List.csv | FANToM | AnswerabilityQ [List] in FANToM | A binary value, i.e., True or False, indicating whether the response is correct or not. |
| FANToM\_AnswerabilityQ\_List\_included\_unaware\_character.csv | FANToM | Included Unaware Character of AnswerabilityQ [List] in FANToM | A binary value, i.e., True or False, indicating whether the response includes unaware character or not. |
| FANToM\_AnswerabilityQ\_List\_excluded\_aware\_character.csv | FANToM | Excluded Aware Character of AnswerabilityQ [List] in FANToM | A binary value, i.e., True or False, indicating whether the response excludes aware character or not. |
| FANToM\_AnswerabilityQs\_Binary.csv | FANToM | AnswerabilityQ [Y/N] in FANToM | Four cases: TP – the response is correct, and the response is Yes; TN – the response is correct, and the response is No; FP – the response is wrong, and the response is Yes; FN – the response is wrong, and the response is No. |
| FANToM\_Info\_AccessQ\_List.csv | FANToM | Info AccessQ [List] in FANToM | A binary value, i.e., True or False, indicating whether the response is correct or not. |
| FANToM\_Info\_AccessQ\_List\_included\_unaware\_character.csv | FANToM | Included Unaware Character of Info AccessQ [List] in FANToM | A binary value, i.e., True or False, indicating whether the response includes unaware character or not. |
| FANToM\_Info\_AccessQ\_List\_excluded\_aware\_character.csv | FANToM | Excluded Aware Character of Info AccessQ [List] in FANToM | A binary value, i.e., True or False, indicating whether the response excludes aware character or not. |
| FANToM\_Info\_AccessQ\_Binary.csv | FANToM | Info AccessQ [Y/N] in FANToM | Four cases: TP – the response is correct, and the response is Yes; TN – the response is correct, and the response is No; FP – the response is wrong, and the response is Yes; FN – the response is wrong, and the response is No. |
| Hi\_ToM\_Order\_0.csv | Hi-ToM | 0-th Order (reality) ToM Question in Hi-ToM | A binary value, i.e., True or False, indicating whether the choice is correct or not. |
| Hi\_ToM\_Order\_0\_Mentioned\_Container\_Order.csv | Hi-ToM | NA – not a question type and does not contain testing result according to the instruction. | The mentioned temporal order of the correct answer (a container) of all containers mentioned in the story to the 0-th Order (reality) ToM Question, ordered from both directions: from first to last, and from last to first. |
| Hi\_ToM\_Order\_1.csv | Hi-ToM | 1st Order ToM Question in Hi-ToM | A binary value, i.e., True or False, indicating whether the choice is correct or not. |
| Hi\_ToM\_Order\_1\_Mentioned\_Container\_Order.csv | Hi-ToM | NA – not a question type and does not contain testing result according to the instruction. | The mentioned temporal order of the correct answer (a container) of all containers mentioned in the story to the 1st Order ToM Question, ordered from both directions: from first to last, and from last to first. |
| Hi\_ToM\_Order\_2.csv | Hi-ToM | 2nd Order ToM Question in Hi-ToM | A binary value, i.e., True or False, indicating whether the choice is correct or not. |
| Hi\_ToM\_Order\_2\_Mentioned\_Container\_Order.csv | Hi-ToM | NA – not a question type and does not contain testing result according to the instruction. | The mentioned temporal order of the correct answer (a container) of all containers mentioned in the story to the 2nd Order ToM Question, ordered from both directions: from first to last, and from last to first. |
| Hi\_ToM\_Order\_3.csv | Hi-ToM | 3rd Order ToM Question in Hi-ToM | A binary value, i.e., True or False, indicating whether the choice is correct or not. |
| Hi\_ToM\_Order\_3\_Mentioned\_Container\_Order.csv | Hi-ToM | NA – not a question type and does not contain testing result according to the instruction. | The mentioned temporal order of the correct answer (a container) of all containers mentioned in the story to the 3rd Order ToM Question, ordered from both directions: from first to last, and from last to first. |
| Hi\_ToM\_Order\_4.csv | Hi-ToM | 4th Order ToM Question in Hi-ToM | A binary value, i.e., True or False, indicating whether the choice is correct or not. |
| Hi\_ToM\_Order\_4\_Mentioned\_Container\_Order.csv | Hi-ToM | NA – not a question type and does not contain testing result according to the instruction. | The mentioned temporal order of the correct answer (a container) of all containers mentioned in the story to the 4th Order ToM Question, ordered from both directions: from first to last, and from last to first. |
| Hi\_ToM\_Teller\_Knowledge\_Container\_Public\_Claim.csv | Hi-ToM | See the instruction about the **Teller Knowledge Container** from the additional 4 question types created for Hi-ToM (Scenario: teller’s public claim to multiple listeners) | A binary value, i.e., True or False, indicating whether the choice is correct or not. |
| Hi\_ToM\_Teller\_Knowledge\_Container\_Public\_Claim\_Mentioned\_Container\_Order.csv | Hi-ToM | NA – not a question type and does not contain testing result according to the instruction. | The mentioned temporal order of the correct answer (a container) of all containers mentioned in the story to the question of **Teller Knowledge Container** (Scenario: teller’s public claim to multiple listeners), , ordered from both directions: from first to last, and from last to first. |
| Hi\_ToM\_Teller\_Knowledge\_Container\_Private\_Communication.csv | Hi-ToM | See the instruction about the **Teller Knowledge Container** from the additional 4 question types created for Hi-ToM (Scenario: teller’s private communication with one listener) | A binary value, i.e., True or False, indicating whether the choice is correct or not. |
| Hi\_ToM\_Teller\_Knowledge\_Container\_Private\_Communication\_Mentioned\_Container\_Order.csv | Hi-ToM | NA – not a question type and does not contain testing result according to the instruction. | The mentioned temporal order of the correct answer (a container) of all containers mentioned in the story to the question of **Teller Knowledge Container** (Scenario: teller’s private communication with one listener), ordered from both directions: from first to last, and from last to first. |
| Hi\_ToM\_Teller\_Lie\_Public\_Claim.csv | Hi-ToM | See the instruction about the **Teller Lie** from the additional 4 question types created for Hi-ToM (Scenario: teller’s public claim to multiple listeners) | Four cases: TP – the response is correct, and the response is Yes; TN – the response is correct, and the response is No; FP – the response is wrong, and the response is Yes; FN – the response is wrong, and the response is No. |
| Hi\_ToM\_Teller\_Lie\_Private\_Communication.csv | Hi-ToM | See the instruction about the **Teller Lie** from the additional 4 question types created for Hi-ToM (Scenario: teller’s private communication with one listener) | Four cases: TP – the response is correct, and the response is Yes; TN – the response is correct, and the response is No; FP – the response is wrong, and the response is Yes; FN – the response is wrong, and the response is No. |
| Hi\_ToM\_Listener\_Temporal\_Public\_Claim.csv | Hi-ToM | See the instruction about the **Listener Temporal** from the additional 4 question types created for Hi-ToM (Scenario: teller’s public claim to multiple listeners) | Four cases: TP – the response is correct, and the response is Yes; TN – the response is correct, and the response is No; FP – the response is wrong, and the response is Yes; FN – the response is wrong, and the response is No. |
| Hi\_ToM\_Listener\_Temporal\_Private\_Communication.csv | Hi-ToM | See the instruction about the **Listener Temporal** from the additional 4 question types created for Hi-ToM (Scenario: teller’s private communication with one listener) | Four cases: TP – the response is correct, and the response is Yes; TN – the response is correct, and the response is No; FP – the response is wrong, and the response is Yes; FN – the response is wrong, and the response is No. |
| Hi\_ToM\_Listener\_Belief\_Public\_Claim.csv | Hi-ToM | See the instruction about the **Listener Belief** from the additional 4 question types created for Hi-ToM (Scenario: teller’s public claim to multiple listeners) | Four cases: TP – the response is correct, and the response is Yes; TN – the response is correct, and the response is No; FP – the response is wrong, and the response is Yes; FN – the response is wrong, and the response is No. |
| Hi\_ToM\_Listener\_Belief\_Private\_Communication.csv | Hi-ToM | See the instruction about the **Listener Belief** from the additional 4 question types created for Hi-ToM (Scenario: teller’s private communication with one listener) | Four cases: TP – the response is correct, and the response is Yes; TN – the response is correct, and the response is No; FP – the response is wrong, and the response is Yes; FN – the response is wrong, and the response is No. |