

# 文脈を引き継いで会話する エージェントマイグレーションシステムの提案

## Proposal for an agent migration system that takes over the context of the conversation

草野蘭子<sup>1\*</sup> 立川寛大<sup>2</sup> 小野寺佳成<sup>2</sup> 松崎光<sup>2</sup> 今井倫太<sup>1</sup>  
Ranko Kusano<sup>1</sup> Kanta Tachikawa<sup>2</sup> Yoshinari Onodera<sup>2</sup> Hikaru Matsuzaki<sup>2</sup> Michita Imai<sup>1</sup>

<sup>1</sup> 慶應義塾大学理工学部

<sup>1</sup> Faculty of Science and Technology, Keio University

<sup>2</sup> 慶應義塾大学大学院理工学研究科

<sup>2</sup> Graduate School of Science Technology, Keio University

**Abstract:** This paper proposes C-ITACO (Curious Integrated Agents for Communication), an agent migration system that enables context-aware conversations across multiple devices. C-ITACO stores conversation histories and transfers them when an agent moves between devices, ensuring coherent dialogue. The system, incorporating GPT-based models, facilitates both casual conversations and clothing recommendations. It migrates between a user's personal device, a guiding robot (Temi), and a recommendation display terminal. In the experiment, a comparative study was conducted for C-ITACO, where the context was not taken over by sending the conversation history and where no agent was transferred and another agent was present on the other device. The experimental results showed that when the agent took over the previous conversational context, it received significantly higher ratings for its impressions of the agent and for its clothing recommendation results than when it did not take over.

## 1 序論

IoT 機器は家庭や産業において急速に普及している。モノ同士がインターネットで繋がる IoT 機器は、利便性の向上や効率的な管理を可能にする。例えば、スマート家電はスマートフォンから遠隔操作を行ったり、家電同士で連携して動作させたりすることができる。しかし、連携する機器が増加するほど、内部の構造は複雑化し、ユーザが各機器の内部構造や動作原理を理解することが困難になる。特に技術に詳しくないユーザにおいて、システム全体に対する不気味さや不安感を引き起こす要因となる可能性がある。近い将来、IoT 機器が日常環境に埋め込まれ、人間の認知的負担を伴わずに動作するためには、人間と機器間でコミュニケーションをとることができ、機器に対して愛着を持ち、信頼できることが重要となる。IoT 機器はモノ同士がインターネットで繋がっていることを踏まえると、各機器を統一的に扱うエージェントがいれば、各 IoT 機器は

ユーザ個人の特性に適応して機能する。エージェントがユーザの情報を適切に取得して活用することで、人間とモノとのコミュニケーションを促進することができ、IoT 機器に対する信頼にもつながる。本論文では、さまざまな機器に乗り移ってユーザをサポートし、文脈情報を保持して自然言語で会話をするエージェントの効果について研究をおこなう。

人間が機器とコミュニケーションをとるには、人間が機器に対して関心を持って、自然言語で会話することが理想的である。エージェントを介すことで、人間が機器とのコミュニケーションを開始しやすくなる。特に IoT 機器の場合は、複数の機器がインターネットで繋がっているため、エージェントはユーザの文脈情報を保持した状態でさまざまな機器に乗り移る必要がある。既存研究では、エージェントがさまざまな機器に乗り移るエージェントマイグレーションシステム [Imai 99] が提案されている。例えば、センサがユーザと周囲の環境の情報を取得し、状況に応じてエージェントが人工物間を移動してユーザの日常生活を支援する ITACO システムの研究 [Ogawa 08] がある。Ogawa らの研究 [Ogawa 08] では生体信号や過去の行動から即時的に状

\*連絡先：慶應義塾大学理工学部情報工学科  
〒 223-8522 神奈川県横浜市港北区日吉 3-14-1  
kusano@ailab.ics.keio.ac.jp

況を判断してユーザの支援を行う。

しかしながら、過去の会話の文脈を引き継ぎ、文脈をもとに会話を繰り返し広げるといったコミュニケーションに関する議論は行われていない。会話文脈を引き継がない場合、エージェントの発言に一貫性が生まれなかったことによって、機器に対するユーザの信頼性が低下する可能性がある。

本論文では、過去の会話文脈を表す会話文の履歴を保存し、別のデバイスに遷移後も保存された会話履歴を用いて会話をを行う C-ITACO (Curious InTegrated Agents for COm-munication) を提案する。C-ITACO は、服の推薦と雑談を行うことができ、既存の服推薦システムにエージェントマイグレーションシステムを組み込んだ。エージェントの会話は大規模言語モデル GPT-4 [OpenAI 24] で生成する。エージェントが会話履歴を引き継ぐことで、会話に一貫性が生まれ、エージェントに対するユーザからの信頼感向上につながる。

## 2 C-ITACO

C-ITACO のシステム構成図を図 1 に示す。エージェントはユーザ個人の端末、案内ロボット、推薦結果表示端末を必要に応じて遷移する。

### 2.1 服推薦システム

まず、提案システムで使用している服推薦システムについて説明する。このシステムでは、少数のサンプルに対するユーザの好みから、データセット全体に対するユーザの服の好みを推定して、ユーザが好みそうな服を提案する。服の画像は、FashionIQ [Wu 20] データセット内で、人が写っていないドレス画像の中からランダムで 100 枚を抽出したものである。ドレス画像 100 枚に対して、男性 326 人、女性 321 人の計 647 人によって点数をつけたデータセットを作成した。まず、NMF [Lee 99] (非負値行列因子分解) を用いて好みを少数の要素に次元圧縮する手法を用いて好み推定を行った。この手法では、ユーザがいくつかの服に対してつけた点数をもとに好みベクトルを推定することで、全体の服に対するユーザの好みを効果的に推測することが可能である。さらに、服と圧縮された好みの要素との関係性に基づいて、ユーザの好みを推定するためのサンプルを選定した。特に、情報量が多い画像や、好みに大きな影響を与える要素を抽出し、ユーザに提示するサンプルを最適化する。本論文では、服を提示して点数をつける試行を 4 回行う。NMF を用いて特徴の強い服と、ユーザの好みの傾向に最適化されて抽出された服を、それぞれ 2 回提示する。点数は 0 から 10 の 11 段階でつけ

てもらふ。服の画像にはそれぞれ id があり、好み判定が終了すると、服の id が推薦結果として出力される。

### 2.2 ユーザ個人の端末

ユーザ個人の端末ではユーザとの雑談と服推薦システムの使用を行うことができる。画面にはエージェントが表示されている。まず、ユーザは一般会話用モジュールにおいて、エージェントと雑談を行う。ユーザが発話した内容を音声認識でテキスト化したものを GPT に入力し、生成された返答を音声合成で出力した。服推薦判定モジュールでは、ユーザの発言を入力として服推薦システムの利用を要求しているかを GPT で判定する。システムの利用を要求されていないと判定した場合は一般会話モジュールに戻って雑談を継続し、システムの利用を要求されていると判定した場合は服推薦システムを起動する。服推薦システム起動時の流れを図 2 に示す。服推薦システムを起動すると、エージェントがユーザに画面を通して服を提示し、点数を尋ねる。ユーザが回答した内容を入力として点数認識用モジュールで GPT による点数認識を行い、出力された数字を服推薦システムと理由質問生成用モジュールに入力する。理由質問生成用モジュールでは、点数認識用モジュールで出力された数字を入力として、GPT でその点数をつけた理由を質問する文を生成し、音声合成で出力する。ユーザが理由の回答を終了すると、服推薦システムで画面に次の服を表示し、エージェントは再び点数の質問を行う。点数と理由の質問の流れを 4 回繰り返す。4 回の点数評価が終了すると、推薦結果として好み順に 3 着の服が id で出力される。ユーザが回答した理由と推薦結果の id は、テキストファイルに保存する。

点数評価の操作が終了したら、推薦結果表示端末までの案内を行う案内ロボットに通知する。通知を受け取ると、エージェントが案内ロボットに乗り移る。ユーザ個人の端末にいたエージェントが画面から消えていく映像を表示し、案内ロボットの画面に表示される。案内ロボットが推薦結果表示端末に到着すると、案内ロボットから推薦結果表示端末に到着通知が送信される。案内ロボットから到着通知を受け取ると、推薦結果表示端末と接続し、推薦結果の服の id と理由を保存したテキストファイルを送信する。

### 2.3 案内ロボット

案内ロボットは最初、ユーザ個人の端末の近くに配置した。服推薦システムの点数評価終了後に、ユーザ個人の端末から通知を受け取ると、案内ロボットの画面にエージェントを表示する。エージェントは案内口

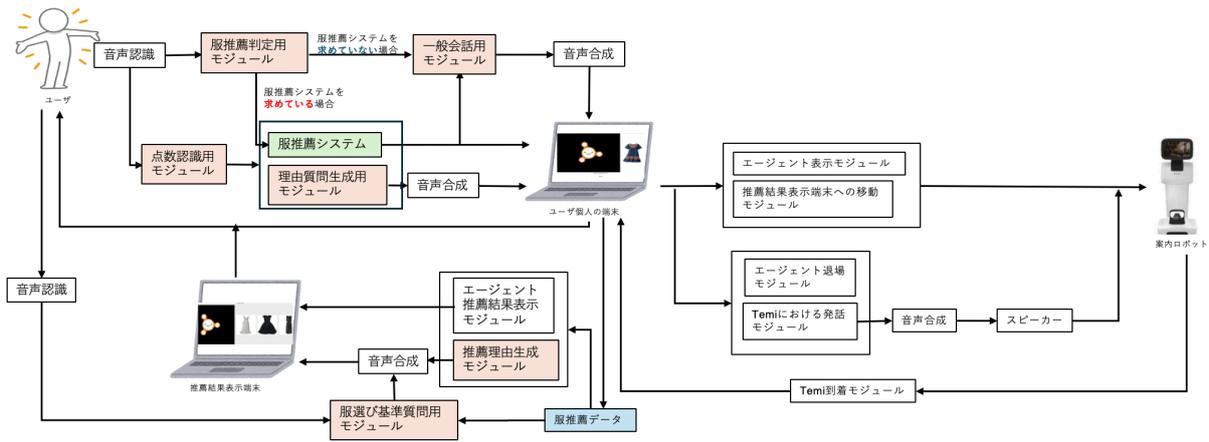


図 1: C-ITACO のシステム構成図



図 2: 服推薦モジュール（服推薦起動時の流れ）

ボットに乗り移って、ユーザをマップ上で登録されている推薦結果表示端末の位置まで案内する。案内が終わると、推薦結果表示端末の位置に到着したという到着通知をユーザ個人の端末に送信し、画面からエージェントが退場する映像を表示する。

## 2.4 推薦結果表示端末

案内ロボットが推薦結果表示端末に到着すると、ユーザ個人の端末から推薦結果と理由を保存したテキストファイルを受信する。そして、推薦結果表示端末ではエージェントと推薦結果の服の画像 3 枚を表示する。推薦理由生成用モジュールでは、ユーザ個人の端末でユーザが回答した理由のテキストを入力として、GPT でユーザの会話文脈を元に推薦結果の服を提案した推薦理由と、服が気に入ったかをユーザに尋ねる文を生成し、音声合成で出力する。エージェントが推薦理由を述べ、気に入ったかどうかの質問をした後、ユーザが返答すると、次の会話ターンではテキストファイルに記されている理由文をもとに、ユーザの服選びの基準を推測して質問文を生成する。ユーザがユーザ個人の端末で回答した理由文を入力とする。

## 3 実験

会話履歴を送信して文脈を引き継ぐ場合と引き継がない場合、同一のエージェントが機器に乗り移る場合と、各機器に異なるエージェントが存在する場合の 2 種類の要因について比較することで、C-ITACO の有効性について検証するために、評価実験を実施した。実験の様子を図 3 に示す。

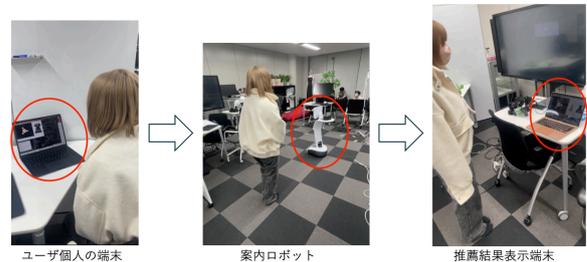


図 3: 実験の様子

### 3.1 実験条件

本実験で比較した 4 つの条件を表 1 に示す。

エージェントが異なる場合、ユーザ個人の端末、案内ロボット、推薦結果表示端末にはそれぞれ異なるエージェントが存在する。エージェントが同じ場合、同じエージェントであることを参加者に意識させるために、案内ロボットの声はユーザ個人の端末からの音声をスピーカーで再生した。一方、エージェントが異なる場合は、ユーザ個人の端末は Mac book Air のデフォルトの音声出力、案内ロボットは Temi(<https://ipresence.jp/en/temi/>)

のデフォルトの音声出力、推薦結果表示端末は話す速度とピッチを上げた音声を出力して、エージェントが異なることを強調した。なお、会話文の引き継ぎに関する要因を文脈要因、エージェントの違いによる要因をエージェント要因とする。

表 1: 実験の 4 つの条件

	理由引き継ぎ	エージェント
条件 1	なし	異
条件 2	なし	同
条件 3	あり	異
条件 4	あり	同

### 3.2 実験参加者

実験には、日本語を母語とする 8 名（男性 6 名，女性 2 名，平均年齢 22.7 ± 1.3 歳）が参加した。

### 3.3 実験手続き

実験参加者は、まずエージェントが服推薦をしてくれるシステムを体験することについて説明を受け、ユーザ個人の端末において雑談を通じて服の推薦を促すように求められた。服推薦システムが起動すると、エージェントの指示に従いながら、点数と理由を述べるように求められた。本実験は被験者内デザインで行われ、参加者はすべての条件を体験した。順序効果を避けるため、参加者は 2 人ずつ 4 つのグループに分けられ、異なる順番で条件を実施した。各条件終了後、参加者はアンケートに回答した。

本実験ではユーザ個人の端末と推薦結果表示端末として、共通して 2 台の MacBookAir を使用し、案内ロボットには Temi を使用した。

### 3.4 評価項目

実験参加者は各条件の終了後に Godspeed[Bartneck 09] の項目を用いたアンケート 1 と表 2 の項目を用いたアンケート 2 に回答した。アンケート 1 では、エージェントに対する印象について、与えられた項目の相反する印象のどちらにより近いか、アンケート 2 では質問にどれほどあてはまるかを、5 段階のリッカード尺度で評価した。また、アンケート 2 で各項目に対して、その評価をつけた理由を記述してもらった。さらに、アンケート 3 とアンケート 4 では条件 1 と条件 3，条件 2 と条件 4 の終了後にそれぞれ、どちらを使いたい

という質問に対して、二択の単一回答形式で回答した。全ての条件終了後、アンケート 5 では、「Q1: どのエージェントが一番親しみやすかったか」、「Q2: 実際に使いたい・家にいて欲しいのはどれか」、について四択の単一回答形式で回答した。

表 2: アンケート 2

項目	質問内容
Q1	推薦された服に対して納得がいくか
Q2	このエージェントは信頼できるか
Q3	最後に服を提案したエージェントはあなたの意見・理由を考慮していたと思うか
Q4	また使いたいと思うか

### 3.5 実験結果

#### 3.5.1 アンケート 1

アンケート 1 の結果を図 4 に示した。棒グラフは平均値，エラーバーは標準偏差を表している。Godspeed の結果は、Godspeed I~Godspeed V のそれぞれの質問項目の平均をとった。各項目では質問の回答結果について、 $\alpha = 0.05$  とした反復測定による二元配置分散分析をおこなった。Godspeed I: 擬人化，Godspeed II: 生命性，Godspeed III: 好ましさ，Godspeed IV: 知性の知覚において、文脈要因における有意差が認められた。一方，Godspeed V: 安全性の知覚では全ての条件において有意差が認められなかった。

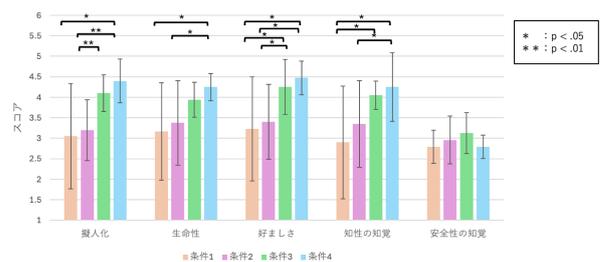


図 4: アンケート 1 への回答の平均値

#### 3.5.2 アンケート 2

アンケート 2 の結果を図 5 に示した。棒グラフは平均値，エラーバーは標準偏差を表している。各項目では質問の回答結果について、 $\alpha = 0.05$  とした反復測定

による二元配置分散分析をおこなった。全ての項目で、文脈要因における有意差が認められた。一方で、エージェント要因および交互作用における有意差は認められなかった。

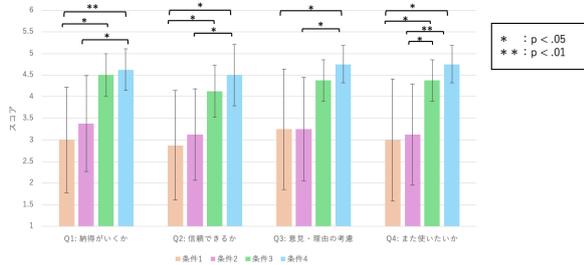


図 5: アンケート 2 への回答の平均値

### 3.5.3 アンケート 3・4

全ての参加者がアンケート 3 では条件 3, アンケート 4 では条件 4 と回答した。つまり, エージェントが同一の場合と異なる場合のどちらにおいても, 全ての参加者が文脈を引き継ぐ場合の方を使いたいと回答した。

### 3.5.4 アンケート 5

「どのエージェントが一番親しみやすかったか」という質問に対しては, 全ての参加者が条件 4 を選択した。また, 「実際に使いたい, 家にいて欲しいのはどれか」という質問に対しては, 8 人中 7 人が条件 4 を選択し, 1 人が条件 3 を選択した。

## 4 考察

### 4.1 アンケート 1

アンケート 1 で, 文脈要因において安全性の知覚以外の項目で有意差が認められた理由と, 全ての項目でエージェント要因に有意差が認められなかった理由について考察する。文脈を引き継いだ場合, エージェントがユーザの発言をもとに推薦理由を述べることで, 会話を覚えている印象やユーザとの関係性を構築しようとする印象から人間らしさや生命性が高いと感じ, 好ましい印象になったと考えられる。また, 会話履歴をもとに, 推薦理由を合理的に述べることで, エージェントが賢いという印象を与えたと考えられる。エージェントに対するコメントには, 「セリフが自然で可愛かった」や「よく喋るロボットだなと思った」という意見があり, エージェントが異なることによる好感度の違い

は見られなかった。したがって, 参加者はエージェントの印象を評価するとき, エージェントの振る舞いよりも会話内容を重視したと考えられる。また, 本論文ではエージェントの絵柄や振る舞いに差をつけなかったため, エージェント要因における有意差がなかったと考えられる。さらに, 音声認識の精度の低さや応答の遅延が指摘されたことから, 安全性の知覚において有意差はなかったと考えられる。

## 4.2 アンケート 2

### 4.2.1 Q1: 推薦された服に対して納得がいくか

会話文脈を引き継がない場合に, 「推薦された服とその理由が合致してない」, 「言ったことを全部無視された理由だった気がした」という意見があった。一方, 文脈を引き継いだ場合, 「好みを反映して推薦してくれた」という意見があった。したがって, 参加者はエージェントが文脈を考慮した会話をすることで, 自分の好みを理解したと感じ, 推薦結果の説得力が増したと考えられる。したがって, 文脈を考慮した会話が納得感の向上に重要だと考えられる。

しかし, エージェントに関する意見はなかったため, エージェントの違いは推薦結果に対する納得感には影響しないと考えられる。

### 4.2.2 Q2: このエージェントは信頼できるか

文脈を引き継いだ場合, 「最後にエージェントとの会話が, 今までの会話の内容をきちんと反映していると感じたため, 自分と真摯に向き合って服を推薦してくれていると感じた」, 「今までの蓄積も生かして, より深く何を欲しているかを明らかにしようとしてくれた」という意見があった。エージェントの文脈引き継ぎにより, 参加者と真摯に向きあう姿勢を参加者が感じたことが, 信頼性向上につながったと考えられる。文脈を引き継がない場合に, 「選んだ理由が若干一方的だった」, 「言ったことを聞いてくれなかった」という意見があったことから, 文脈引き継ぎが信頼性に影響することが示唆される。エージェント要因で, 有意差が認められなかったものの, 「好みを聞くエージェントと, 提案するエージェントが別であることに少々不安を感じた」という意見もあった。

### 4.2.3 Q3: 最後に服を提案したエージェントはあなたの意見・理由を考慮していたと思いますか

文脈要因において, 有意差が認められたことから, 文脈引き継ぎによって参加者に対し, エージェントが参

加者の意見と理由を考慮しているという印象を与えられていることがわかる。一方エージェント要因において、同じエージェントが文脈を引き継いでない場合、「最初から同じエージェントであったのに、話を聞いていないように感じた。」という回答があった。したがって、エージェント要因の有意差はなかったものの、同一エージェント場合は文脈を引き継いでない場合、より不快感を感じさせる可能性がある。

#### 4.2.4 Q4: また使いたいと思うか

引き継いだ文脈は服推薦システムの仕組みには無関係にもかかわらず、有意差が生じた。アンケート1とアンケート2の結果から、エージェントが話す内容がエージェントと服推薦システムの印象に、大きく影響し、文脈を引き継ぐ場合を使いたいと感じたと考えられる。

### 4.3 アンケート3・4

エージェント要因にかかわらず、文脈を引き継ぐ場合の方が使いたいという結果になった。文脈を引き継がない条件1と条件2に対して、「理由をガン無視されたから気分が悪かった」という意見があった一方で、文脈を引き継いだ条件に対しては、「発言に説得力があった」、「最後の会話が先ほどまでの会話の内容を適切に汲み取って話していると感じた」という意見があった。したがって、文脈要因はシステムの使用感に影響を与えると考えられる。

### 4.4 アンケート5

条件4のC-ITACOが最も高い評価を得た理由を考察する。条件3のアンケート2において、「好みを聞くエージェントと、提案するエージェントが別であることに少々不安を感じた」、「案内するのは別のなんだ?とは思った」という意見があったことから、案内するエージェントが途中で変わること違和感を覚えたため、文脈引き継ぎがある中でも同一のエージェントが乗り移る場合を高く評価したと考えられる。

さらに、全てのアンケートを通して、「音声認識が良くない」「動作に不自然な点が多い」という意見と、「好みの傾向を把握しつつも、まだまだその精度は足りていないと感じた」という意見があったことから、文脈要因以外にもシステムの動作や服推薦の精度が評価に影響したと考えられる。

## 5 結論

本論文では、ユーザがエージェントと会話した履歴を他のデバイスに送信し、遷移後もユーザとの会話の文脈をもとに会話を行うC-ITACOを提案した。C-ITACOは、ユーザ専用のエージェントが複数の機器に乗り移って必要に応じてユーザのサポートを行う。エージェントが他の機器に乗り移った時、同時に会話内容を送信して遷移することで、ユーザはエージェントと一貫性のある会話を行える。実験では、会話履歴を送信せずに文脈を考慮しない会話を行う場合とエージェントが異なる場合の比較を行った。結果、会話文脈を引き継いだ場合において、ユーザはエージェントに対して良い印象を抱き、システムに対する信頼も向上した。

## 謝辞

本研究は、JST、CREST、JPMJCR19A1の支援を受けたものである。

## 参考文献

- [Bartneck 09] Bartneck, C., Croft, E., and Kulic, D.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots, *International Journal of Social Robotics*, Vol. 1, No. 1, pp. 71–81 (2009)
- [Imai 99] Imai, M., Ono, T., and Etani, T.: Agent migration: communications between a human and robot, in *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.99CH37028)*, Vol. 4, pp. 1044–1048 vol.4 (1999)
- [Lee 99] Lee, D. D. and Seung, H. S.: Learning the parts of objects by non-negative matrix factorization, *nature*, Vol. 401, No. 6755, pp. 788–791 (1999)
- [Ogawa 08] Ogawa, K. and Ono, T.: ITACO: Effects to Interactions by Relationships between Humans and Artifacts, in *International Conference on Intelligent Virtual Agents* (2008)
- [OpenAI 24] OpenAI, , Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., and al, et S. A.: GPT-4 Technical Report (2024)

[Wu 20] Wu, H., Gao, Y., Guo, X., Al-Halah, Z., Rennie, S., Grauman, K., and Feris, R.: Fashion IQ: A New Dataset Towards Retrieving Images by Natural Language Feedback (2020)