AutoChemplete - Making Chemical Structural Formulas Accessible

Merlin Knaeble
Gabriel Sailer
Zihan Chen
Thorsten Schwarz
Kailun Yang
Karlsruhe Institute of Technology
Germany

Mario Nadj
TU Dortmund University
Germany

Rainer Stiefelhagen
Alexander Maedche
Karlsruhe Institute of Technology
Germany

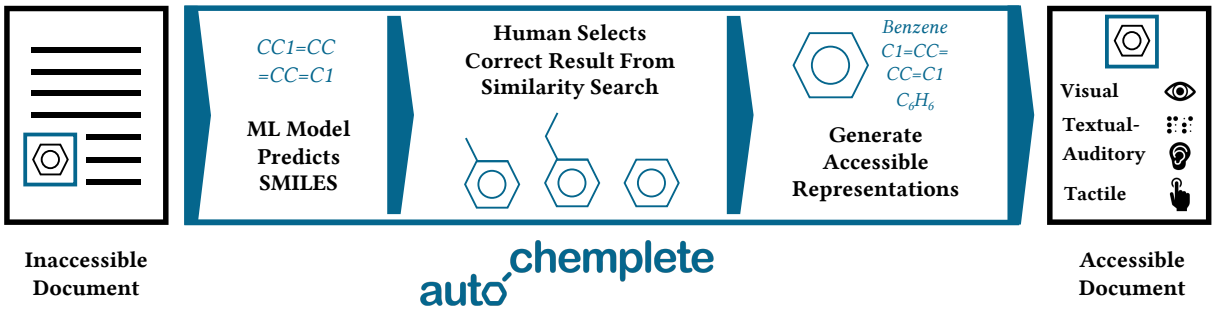


Figure 1: Processing pipeline of *AutoChemplete*, ingesting a bitmap image of a chemical structural formula from an inaccessible document, using interactive labeling to create various accessible visual, textual-auditory, or tactile representations.

ABSTRACT

Despite their interests, blind and low vision students avoid STEM subjects. Research attributes this to a lack of accessible material. Annotating STEM content such as chemical structural formulas requires expert domain knowledge, is time consuming, and frustrating. We conduct interviews with blind and low vision chemists and accessibility professionals to derive requirements for tool support. On this basis, we develop AutoChemplete, an interactive labeling tool for chemical structural formulas. It ingests images and uses machine learning to predict the molecule. With a similarity search in the solution space, we enable even novices to simply pick from options. From this we are able to generate accessible representations. We conduct fifteen think-aloud sessions with participants of varying domain expertise and find support of different annotation styles simultaneously. Not only does AutoChemplete strike a balance in skill-support, participants even find it entertaining.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility systems and tools**; *User studies*; *User interface design.*

KEYWORDS

accessibility, STEM, chemistry, structural formula, interactive labeling, SMILES, autocomplete

ACM Reference Format:
Merlin Knaeble, Gabriel Sailer, Zihan Chen, Thorsten Schwarz, Kailun Yang, Mario Nadj, Rainer Stiefelhagen, and Alexander Maedche. 2023. AutoChemplete - Making Chemical Structural Formulas Accessible. In *20th International Web for All Conference (W4A '23), April 30-May 1, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3587281.3587293

W4A '23, April 30-May 1, 2023, Austin, TX, USA

ACM ISBN 979-8-4007-0748-3/23/04…$15.00
https://doi.org/10.1145/3587281.3587293

1 INTRODUCTION

Science, technology, engineering, and mathematics (STEM) education lacks accessibility for blind and low vision (BLV[1]) students. Among US students with vision impairments, three-quarters are *"more than a full grade level behind their sighted peers in mathematics"* [20]. Although 69% of US BLV students show interest in STEM during high school, only 8% pursue related college degrees [50]. A key reason why BLV students do not choose STEM subjects in higher education is a lack of accessible learning material [24]. In many STEM fields, visual elements like diagrams regularly convey important information. For example, without structural formulas visualizing atoms and bonds, studying chemistry seems nearly impossible [27, 35].

[1]We follow the World Health Organizations (WHO) definitions of low vision (performing visual tasks at reduced level, with less than 30% vision remaining with the best possible correction), and blind (vision unavailable or unreliable, with less than 5% remaining after correction). For details see https://id.who.int/icd/entity/1103667651

Research, driven in considerable parts by the HCI, web, and accessibility communities, led to improvements, especially for digital document access [33]. Tactile displays can show matrices and chemical formulas in LaTeX or as vector graphics [38], while auditory descriptions are used for math formulas [20]. More recent works support BLV authors in collaborative editing [13, 31].

However, STEM content is often provided in form of PDFs, for instance textbooks, assignment sheets, and papers, many of which do not generally meet accessibility requirements [16]. Even with comparably mild impairments regarding color vision, about 40% of figures in publications are rated as inaccessible [4]. 85% of US BLV students in STEM report to receive important course material later than their peers, as it needs to be made accessible first [8]. This process has repeatedly proven to be costly and time-consuming [38], as well as, frustrating and error-prone in other labeling contexts [37]. For tables in HTML documents, automated approaches have shown promising results as the original data is already in a structured form and simply needs to be re-processed [55]. However, for widely used PDF documents, such fully-automated approaches fail, specifically when presented with those very structural elements in the form of graphics of importance for STEM education [35, 48]. Where manual approaches struggle to scale and automation fails regarding the required exactness, an interactive approach combining the strengths of humans and computers seems adamant. Hereby, AI-based extraction often delivers a useful first solution, to further iterate upon manually. To this end, a well-known feature for text entry on mobile phone keyboards [44] or web search queries [40] is autocomplete. In this, the computer offers multiple potential continuations for a given entry. Such features have already proven beneficial in scenarios where novices are asked to complete tasks usually better performed by experts [32].

To address the problem of provisioning accessible annotations of chemical structural formulas, we introduce *AutoChemplete*, a novel interactive labeling tool. Figure 1 offers a high-level overview. We designed *AutoChemplete* to take over after extracting an image of a structural chemical formula from a document. We then automatically extract a first prediction of a so-called SMILES[2] [56] string with an encoder-decoder model. SMILES is a unique linear representation of molecules that is both machine- and human-readable. Based on this, we perform a similarity search on the entirety of the practically possible solution space (i.e. a database of known molecules) to obtain an ordered list of potential matches. After human approval or correction, we can generate alternate representations, including visual ones for sighted readers, linearized descriptions like name, IUPAC[3] [19], or SMILES, as well as tactile output, like vector graphics with embedded Braille.

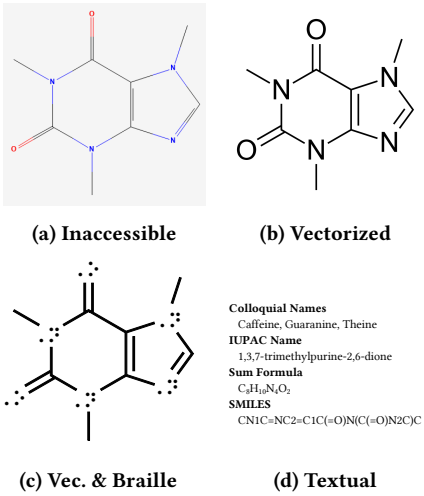
In short, aside from developing *AutoChemplete*, we contribute with requirements for alt text tool support, as well as a user evaluation. Further, we deliver a concept to generalize beyond our use case. In the following, we first present related work. Then, we outline the design process for *AutoChemplete*. We conducted interviews with three BLV chemists and four accessibility professionals to elicit requirements for tool support. On this basis, we present *AutoChemplete*. Further, we quantitatively evaluated the performance our ML model and qualitatively evaluated *AutoChemplete* in think-aloud sessions with 15 participants of different backgrounds, ranging from chemistry professor to students of subjects well outside STEM. Finally, we discuss our findings, state their limitations, as well as avenues for future research, and conclude our paper.

[2]The Simplified Molecular-Input Line-Entry System is a line notation for describing the chemical structure by means of short strings.

[3]The International Union of Pure and Applied Chemistry defines a set of rules to assign names to molecules, which represent their structure.

2 RELATED WORK

In this section, we provide an overview of research on STEM document accessibility, automated recognition for chemical formulas, as well as interactive labeling and autocomplete in different domains.



(a) Inaccessible (b) Vectorized

(c) Vec. & Braille (d) Textual

Figure 2: Depiction of caffeine from https://pubchem.ncbi.nlm.nih.gov/compound/2519 and accessible representations

2.1 STEM Document Accessibility

Figure 2a shows a depiction of caffeine from the widely used chemical database PubChem [26], which shares common issues for people with color blindness and other vision impairments, like color contrast and line thickness. Accessibility and HCI research has produced several ways for BLV chemists to access such depictions. Linear representations (Figure 2d), like IUPAC names or the simplified molecular-input line-entry system (SMILES[4]) are easily compatible with screen readers or Braille output. Vectorized representations, either with Latin (Figure 2b) or Braille (Figure 2c) lettering allows for necessary contrast, coloring, line spacing and thickness, as well as magnification. Potential non-visual output modalities that keep the two-dimensional aspect include tactile displays, swell and embossed paper, or semantically enriched vector graphics for digital exploration [24]. Education of BLV students has been a focus for makers, evaluating 3D printing and laser cutting to produce accessible learning material ad-hoc [10]. Koone et al. [30] have recently made biochemistry plots accessible in form of 3D prints. Diagrams, frequently used to visualize measurement data, have been made

[4]Traditional SMILES itself is not unique, i.e. multiple strings can represent one molecule. Chemists have developed methods to compute so-called canonical SMILES strings from any SMILES, which are unique. We always refer to canonical SMILES.

accessible by combining sonofication and voice [23]. Maps, another typical case of 2D information needing to be made accessible [22], have been successfully converted into tactile formats automatically [21]. However this requires annotated data. Where such data is not available, for instance for scene descriptions of video content, research has successfully relied on human-in-the-loop ML [59]. Gamification helped to recruit volunteers for producing alt texts [54]. Overall, though, research focusing on the procurement of accessible content [29, 54, 59] seems to be the exception to the rule, as scholars' focus is rather on how BLV users can consume such content [e.g. 10, 23, 30]. Research on accessible chemistry education for BLV students by D'Agostino [18] summarizes these needs by calling for *"digital textbooks and [...] education resources that are compatible with accessibility standards"*, as well as providing *"tactile [...] representations"* and *"alt text descriptive annotations"*.

2.2 Automated Recognition of Chemical Structural Formulas

Automated recognition of structural formulas is also of interest outside of accessibility concerns. Research driven by chemists, as well as ML experts has consequently made ample efforts to improve this process. Early on, rule-based approaches were the norm. Researchers [34, 41, 47, 52] regularly combined simple image processing approaches like vectorization with rules along which they could recognize molecules. Such rule-based approaches are often highly sensitive to noise, additionally they rarely generalize knowledge by themselves, and must rather be tuned and optimized manually [12]. More modern approaches use deep learning, especially Transformer-based encoder-decoder models [43, 49, 58]. Most recently [58] report accuracies of $67-79\%$, while [43] report SMILES generation accuracies of $37-87\%$. [49] report accuracies of $77-83\%$.

2.3 Interactive Labeling & Autocomplete

Qian *et al.* [43] acknowledge the need for manual correction of ML-predicted SMILES, as their model still is not fully reliable. To this end, they conduct a study with a single chemistry student. They show that molecular graphs help experts to more efficiently annotate SMILES to images. On this basis, they report a 50% decrease in annotation time. Interactive labeling approaches in other contexts have reported similar benefits when combining human and artificial intelligence [61]. Chemists have developed *kekule.js*, an interactive editor for structural formulas. It supports graphical entry and can output SMILES, without the user knowing this representation [25]. HCI research has made similar advances for mathematics, commonly denoted by LaTeX math-mode. *MathDeck* [17] allows for live rendering in a handwriting-like format, manipulation without knowledge of the linearized format, and search for similar content. As such it allows for a rudimentary autocomplete implementation for math formulas. Autocomplete has a long history in HCI research and is best known from web search [40] or as predictive text entry on touch keyboards [44]. In the context of breadboard circuits, autocomplete features have recently shown to support novices in becoming as effective as their expert peers [32]. Human-AI collaborative approaches, in form of interactive labeling, have further shown to improve efficiency and effectiveness of comparable data entry tasks in various contexts [e.g. 11, 15, 60].

3 DESIGN OF AUTOCHEMPLETE

We designed *AutoChemplete*[5] as part of a larger (unreleased) document accessibility platform that takes over tasks like reading order annotation and identification of content elements. Such general functionality is also provided by commercial tools like Adobe Acrobat Pro[6], or free ones like PAVE[7], to which *AutoChemplete* is complementary. *AutoChemplete* was designed for real-world usage at the center for accessibility of our university. One of the main tasks of the center is to provide accessible literature and learning materials to all enrolled BLV students. As such, the employees of the center make teaching material for more than 120 courses per semester accessible. This number is expected to grow significantly, increasing the need for tool support (i.e. in form of a dedicated document accessibility platform) and at least partial automation.

3.1 Interviews

As a starting point for designing *AutoChemplete* we conducted semi-structured interviews with two groups. First, we were interested in the perspectives of BLV chemists. BLV chemists are not users of *AutoChemplete* itself. However they are dependent on the product of *AutoChemplete*, accessible documents with chemical structural formulas. Of the five BLV chemists we contacted, we were able to interview three. The others supplied their vision impairment, education, job, and currently most used molecule representation via email. We summarize this in Table 1. Second, we recruited four employees of the center for accessibility to gain insights into the annotation process, rather than the final product. This group are among the target users for *AutoChemplete*, as it regrettably requires a sighted person to convert an inaccessible PDF into an accessible one. We choose accessibility professionals, as they have ample experience. Participation was voluntary and no compensation was offered. Of our seven participants one reported her gender as female. Ages ranged between 21 and 45 years, with an average of 29 (SD=9).

Interviews were conducted via online (video) meetings and took on average 21 minutes for A1-4 and 63 minutes for C1-3. The interviews were recorded and transcribed verbatim. BLV chemists were queried about their access to literature, preferred representations for molecules, and challenges around STEM education. Accessibility professionals were asked to report on the current status quo and challenges around making literature accessible. To deduct requirements we conducted a thematic analysis following Braun and Clarke [9]. For efficiency, coding was performed on a one-interview-one-coder basis. The two coders then jointly performed aggregation steps. Across the five interviews we obtained a total of 575 in-vivo codings which we first aggregated into 70 code groups. From these we identified five actionable requirements, while 11 code groups remained as a more overarching goal regarding the lack of accessibility and willingness for inclusion in STEM education[8].

[5]We release AutoChemplete as open source as a standalone application: https://github.com/human-centered-systems-lab/AutoChemplete
[6]https://www.adobe.com/acrobat/
[7]https://pave-pdf.org/
[8]We offer guidelines for both groups and a codebook mapping code groups to requirements in the supplemental. We began with in-vivo coding selecting passages as-is, before merging into code groups and finally requirements (as themes).

ID	vision	chem. edu.	occup.	representation	3^{rd}
C1	blind	MSc	industry	tactile paper	✓
C2	blind	HS	B.Sc.	tactile magnets	✓
C3	low vis.	MSc	Ph.D.	zoom & colors	-
-	blind	PhD	industry	SMILES (Braille)	(✓)
-	blind	PhD	PostDoc	tactile display	✓
ID	**edu.**	**occupation**			
A1	PhD phy.	lead of literature team at a11y center			
A2	MSc CS	accessibility researcher			
A3	BSc CS	MSc & document accessibility platform dev.			
A4	BSc CS	MSc & document accessibility platform dev.			

Table 1: Overview of the requirements interviewees: BLV chemists and accessibility professionals. Note that two BLV chemists could not participate, but provided information on their preferred molecule representations. All BLV chemists also consume e.g. names and IUPAC convention via screen readers, except for C3, who is able to read with their eyes. 3^{rd} represents whether their current molecule representation requires a third party to assist in transferring it into this format every time the molecule is to be read.

3.2 Requirements

Based on the interviews, we can confirm a severe lack of accessibility of STEM learning material various prior works have identified [e.g. 8, 24, 50]. Interviewees report that if there is accessible learning material it needed other BLV students before them *"that championed accessibility issues in the student representation"* (C2). Others state to be *"more or less the first"* (C1) BLV chemist at their respective university. One participant recalled that the representative for people with disabilities at a university they applied to asked why *"they just won't consider doing something else, like studying psychology, going to law school, or becoming a manual therapist"* (C2). Another participant was suggested to *"switch majors from chemistry to any language, to avoid such [accessibility] problems"* (C3).

R1 - Exact Annotations. *"In normal text, we have a lot of redundancy, so a typo won't really affect the understanding"* (A3). Many STEM notations like math formulas or molecular structures are sensitive to mistakes, as they are *"made to be compact"* (A3). Here, *"slight mistakes can make a huge difference, and you don't want BLV students to have less of a chance of learning just because the annotation is wrong"* (A4). The center for accessibility regularly produces *"material for exams"*, where they need *"100% accuracy, so a human check has to always be there"* (A2), as fully-automated approaches do not produce exact results [43]. Hence, *AutoChemplete* should support as much as possible the exactness of the annotations.

R2 - Reasonably Fast. *"Blind chemistry students are in short supply"* (C2), but just one student can (and should) voice the need to have all materials accessible. This is often seen as unsolicited *"extra work"* (C1) by educators. Some universities support literature supply. Of the interviewed BLV chemists, only one has *"enjoyed such luxury"* (C2). Accessibility centers experience fluctuation in demand, annotating *"between five and ten documents per month"* (A2). As each document *"takes quite a long while"* (A1) to annotate, and the process is performed twice for exactness, *AutoChemplete* should be as fast as possible, while not sacrificing R1, by employing autocomplete to more quickly find the correct solution.

R3 - Skill Support. At the accessibility center, annotation is often performed by student assistants. However, availability varies between subjects. *"Especially for chemistry it's difficult, right now its easy for physics, computer science, or maths"* (A1). *"In some cases it's very difficult to describe [a graphic] if you don't understand all of it"*, so *"we need some kind of experience in the field"* (A2), otherwise *"you might not be able to annotate correctly"* (A3). However, as demand fluctuates (R2), accessibility professionals raise the question of *"how physics or math or any non-chemistry student can support those [BLV chemistry] students"* (A1). BLV chemists without support from paid assistants regularly have to rely on friends and family. However, often only people from related subjects *"can do this sufficiently well"*, and supporters with other backgrounds are *"totally unable to cope"* (C3). *AutoChemplete* should, thus, support less skilled users, by offering autocomplete to confirm correctness.

R4 - Output Formats. *"Our process is quite complex, because we have different situations [...] we customize the output for each student"* (A2). Sometimes *"a description is not enough, so we have to do a tactile graphic"* (A2). Table 1 shows how all five BLV chemists choose a different representation. C1 and C2 intend to switch to a digital modality, however C2 lamented the lack of content. High-abstraction formats like sum formulas are an important *"foundation to grasp size and composition of a molecule"* (C2). As *"one is so much slower in a PDF [than a sighted person]"* (C2), BLV chemists need different formats for different needs. Sum formulas and names allow for a quick grasp. If the chemist already knows the structure, they don't need further detail. IUPAC offers first structural hints as it is assembled according to known rules. SMILES offers all details, but needs training to be read. Vectors allow for magnification, color correction, and tactile output. A one-size-fits-all *"solution is impossible"* (C3). Hence, *AutoChemplete* should support a wide range of outputs, however, as voiced by A2, without redundant work.

R5 - Tool Integration. *"What would be nice is wizard guiding you"* (A1). Professionals see the need to be supported by specialized *"tools for [...] visual elements like charts, diagrams, math equations, or chemistry"* (A1, c.f. R3). Current approaches already do *"auto-segmentation for content blocks, [...] and reading order"* (A4). For certain content blocks, like chemical structural formulas, *"one needs to manually add representations"* (A4). Afterwards a *"compliant document"* is exported (A3). On this basis, *AutoChemplete* needs to integrate into existing processes that take over steps like ingestion, OCR, segmentation, reading order, and output.

3.3 Molecule Recognition

To train our model, we sampled 2 million images from the PubChem[9] [26] database, 900k images from DACON[10], another 900k from Bristol-Myers Squibb[11] including heavy noise, which may also occur in the real world, as well as 2 million compounds from

[9] https://pubchem.ncbi.nlm.nih.gov
[10] https://dacon.io/competitions/official/235640/overview/description
[11] https://www.kaggle.com/competitions/bms-molecular-translation/overview

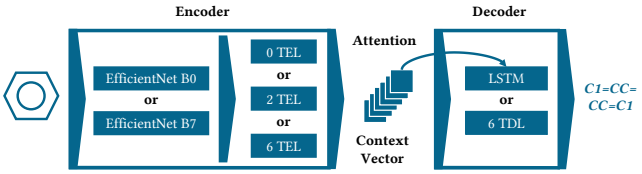


Figure 3: Architecture of our sub-models. The encoder consists of an EfficientNet B0 or B7, and no, two or six Transformer encoder layers (TEL). The decoder is either an LSTM with attention or a six layered Transformer (TDL).

ChEMBL[12] [36], which have a distinct visualization form for bioactive molecules. Over all these data-sets, we use a 80/20 split for training and validation. The images are, if required, downsampled to 256×256 pixels in grayscale. As we find vastly different visualization styles between data sources (c.f. Figure 2), we employ a multi-model ensemble strategy to be able to capture such differences. In each sub-model, we use one of the established pre-trained EfficientNet-B0 or B7 [51] models with or without additional transformer layers [53] to extract image features and generate a context vector. LSTM with attention [5] or transformer decoder layers are then used to generate the SMILES strings. We have trained a total of five sub-models to construct our ensemble. Figure 3 shows an overview of this architecture. Hereby we constructed sub-model 1 and 2 as simple, general purpose versions. We expect them to perform well on clean images, e.g. in uncompressed native PDFs. For sub-models 3 and 4 we introduce noise into the training data, as it can be found in scanned PDFs or as the result of compression artifacts or low-quality images. Finally, we optimize sub-model 5 towards the distinct visualization in ChEMBL. As chemical structural formulas are visualized in a range of styles, our multi-model architecture is individually tunable to such cases. For prediction, we follow a majority/similarity vote of the five sub-models to determine one SMILES. While each sub-model acts independently during generation, we automatically validate syntactical correctness before feeding the strings into the ensemble voting mechanism. RDKit[13] performs this syntax check much like a compiler does for code, trying to construct a molecule based on the SMILES string. It detects syntax errors if, e.g. an atom is assigned more bonds than chemically possible. We maximize the probability of syntactically valid SMILES being produced, as if one sub-model fails to do so, we have others to rely on, as our proposed similarity search requires a valid string. Such ensemble-based architectures have the additional advantage of being easier to train and re-train. And as *AutoChemplete* is designed as an interactive labeling tool, we can incorporate the resulting user-annotated images as part of additional training data, once the system is in productive use. All sub-models have attention and embedding dimensions of 512, without freezing layers during training, as EfficientNet is generally pre-trained for natural images (i.e. photos) as opposed to synthetic structures like our molecular formulas. Hereby we fine-tune the EfficientNet part of the encoder, while training the decoder from scratch. As progress of research on automated recognition of chemical structural formulas is ever so steady, we designed *AutoChemplete* to easily incorporate future, even better performing models.

[12]https://www.ebi.ac.uk/chembl/
[13]https://www.rdkit.org/

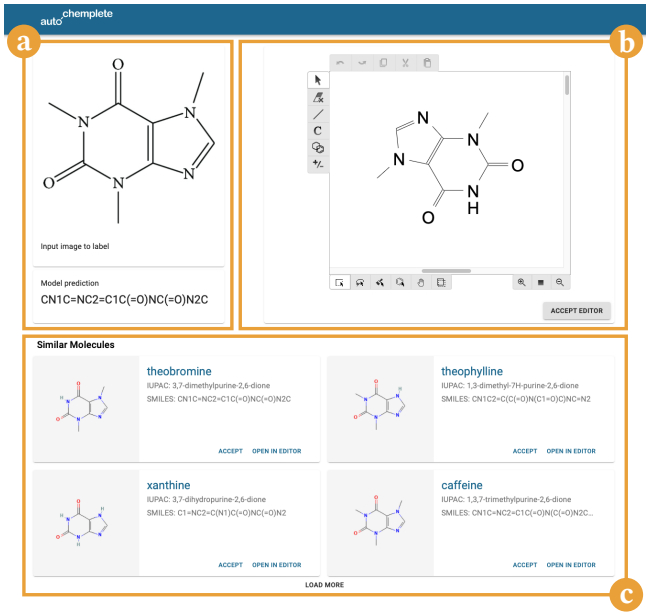


Figure 4: User interface of *AutoChemplete*, with its core components highlighted. Input image from [39].

3.4 User Interface

On the basis of the requirements, we developed *AutoChemplete* as an interactive labeling tool, hence a complementary ML with a human-in-the-loop. Addressing requirement R1 through R3, we designed an interaction around three synergistic components. Our ML model (the first component, see 3.3) takes the input image, and produces an initial suggestion for a SMILES string, as represented in Figure 4a, addressing R2. Then, *AutoChemplete* performs a similarity search in the PubChem database[14] [26]. This search operates on established practices and distance measures. Chemistry generally has agreed upon Tanimoto similarity, based on $ECFP_{6,1024}$ molecular fingerprints for search purposes [12]. String-based distance measures are inferior in capturing molecular differences as opposed to fingerprint-based ones [6]. Further Tanimoto is also established as the benchmark metric for molecular recognition [12, 57, 58]. Such fingerprints offer a unique, albeit as compared to SMILES not human-readable, identifier of molecules, so that chemically related molecules also have close fingerprints. In chemistry, this is helpful to determine similar reactive properties. As $ECFP_{6,1024}$ fingerprints are bit arrays of length 1024, the calculation of the Tanimoto similarity T of two fingerprints A and B becomes computationally efficient at $T(A,B) = \frac{|A \wedge B|}{|A \vee B|}$, corresponding to simple bit-wise comparison. Hence, such similarity values are computed for the molecules in the database, and the results are ranked in descending order. We operate under the reasonable assumption, that molecules displayed in STEM learning material will not be random arrangements of

[14]https://pubchem.ncbi.nlm.nih.gov

bonds and atoms, but rather ones existing in reality. The PubChem database is the largest collection of known molecules, and is as such highly likely to contain the structure in question, or at least a very similar one. The results are ordered by rank and shown as suggestions, see Figure 4c. They form the second component. Initially, we show the top four suggestions to not overload the user interface, however there are further results available on demand. For each suggestion we show information like colloquial names and the IUPAC convention. Further, we present a preview visualization of each suggestion. By clicking on this preview, the user can open a side-by-side comparison.If the user finds the correct molecule in these suggestions, they can click on the accept-button, to directly confirm the corresponding molecule as the correct one (R2, R3). Such suggestions further support R1, the annotations' exactness, as users can compare their annotations to the suggestions. Figure 4b shows the third component of *AutoChemplete*, the editor. Hereby, with *kekule.js*[25], we incorporate a well known open source editor for structural formulas. Relying on an industry standard has the benefit that if users do have domain expertise, they may already be familiar with this part of *AutoChemplete*. Upon invocation the editor comes pre-loaded with the initial model prediction. The user can then accept what is displayed in the editor (R2), or first make manual changes (R1). If the user finds a close, but not correct, molecule in the suggestions (Figure 4c), they can also open it in the editor to make further changes (R2, R3). Additionally, after each manual change, the similarity search is refreshed. This allows for the autocomplete aspect of *AutoChemplete*, where users do not have to reproduce the entire molecule by hand, but can rather rely on AI to finish their work. As such, *AutoChemplete* employs global (as opposed to local) autocomplete as it generally suggests entire solutions instead of partial ones. Our manual editor further offers local autocomplete (e.g. filling in hydrogens).

To support R5, *AutoChemplete* is designed to be used when a structural formula has been found in a document. It takes, as its only input, an image of the molecule. To comply with R4, we deliver all textual representations mentioned in Figure 2c (colloquial name, IUPAC name, sum formula, SMILES). Such textual representations can be accessed via screen reader or automatically converted to Braille, for use on displays or on paper. Additionally we can generate non-human readable formats like InChI[15] and ChemFig[16] for export, as well as a short text containing the number of atoms and bonds, occurring elements, and the molar mass. Once the correct SMILES string is known, such formats are easily retrievable from chemical databases - in our case PubChem, as it is used for the search anyways. Further we can export newly rendered configurable vector graphics for either visual or tactile output. Depending on user preference, for vector graphics, we can keep Latin lettering (e.g. for visually impaired chemists with remaining vision), or replace letters with their Braille equivalent for better tactile output.

4 EVALUATION

We evaluate *AutoChemplete* from two perspectives: first the performance of the underlying ML model, second, with a comprehensive user study, qualitative themes, as well as annotation paths.

[15]The International Chemical Identifier is a linear notation for chemical substances.
[16]ChemFig is a LaTeX package for drawing molecules.

4.1 Molecule Recognition Evaluation

Testing against previously unseen test data, we report two important metrics. As common in most ML applications, we employ an exact match accuracy. Hereby, the model-predicted SMILES is correct, if it exactly corresponds to the unique ground-truth SMILES. Our model achieves an exact match accuracy of 83.36%. Thereby, it surpasses existing works [58] at $67-79\%$ and [49] at $77-83\%$ in accuracy. It places at the higher end of [43] at $37-87\%$, whose accuracy is highly dependent on visualization and molecule distribution. Additionally, we study the overall similarity to the correct result. Our model reaches an average Tanimoto similarity (computed as explained beforehand) of 91.71% surpassing [58] at 86.6%.

We further combine our output with the similarity search provided in *AutoChemplete* to study how quickly the correct molecule can be found in the suggestions. Hereby, we perform such a search based on the prediction and report the rank of the ground truth, calculated as the number of hits with a Tanimoto similarity greater or equal to that of the ground truth, plus one.

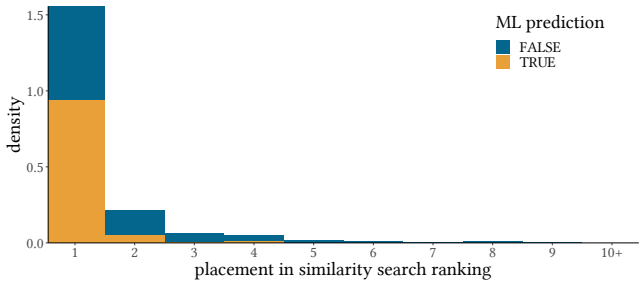


Figure 5: Similarity search ranks for molecules the ML model predicted correctly (yellow) and incorrectly (blue).

To confirm whether our similarity search works, we check for the average similarity search placement of the ground truth within these 83.36% where the model delivers the correct results with an exact match. The corresponding molecule has a mean placement of 1.09, with a median of 1 within the similarity search ranking. Our distribution shows a worst-case of place 4, meaning that for 83.36% of molecules, *AutoChemplete* is able to show the correct molecule including name and IUPAC within those suggestions shown in the interface. This uncertainty in the placement is explained by the input images having different information content regarding stereochemistry and chirality. Some visualizations omit e.g. stereochemical information about the three-dimensionality of molecules. If then different three-dimensional versions of the molecule exist (e.g. D- and L-glucose), the search for the representation without this information returns all. For the remaining 16.64% of molecules, which the model does not identify correctly, we still find a median placement of 1. The mean, however, is heavily skewed to 52.88, by a few outlier molecules where the similarity search can only find the correct match after a worst-case of 16762 others. However, for 62.5% of the molecules that were not already identified by the model, the correct solution remains in the top four suggestions shown. These results are also visualized in Figure 5, showing a long tail. In summary, combining the model with a similarity search we allow for an efficient retrieval of the correct solution. In 93.76% of all

molecules in our test data set we can find the correct solution in the top four suggestions shown in the user interface of *AutoChemplete*.

4.2 User Study

To evaluate *AutoChemplete*, we conducted a user study. We were interested in how the system supports differently skilled users (c.f. Table 2) and whether participants were able to use our tool.

Participants. We recruited a total of 15 participants with chemical expertise varying between a music student and a chemistry professor, and accessibility knowledge of a comparable range. Eight participants were part-time employed as student assistants for literature annotation. As there is currently a high need for accessible learning materials from physics, within those eight there is a skew of the educational backgrounds towards this subject. However, such a sample is of interest for our investigation, as physics students should have a general affinity towards STEM, without specific knowledge in chemistry. We further recruited five experts on the topic of chemistry, as well as two students from subjects outside of STEM. Six participants self-reported as female, the other nine as male. On average, participants were 26 years old (SD=9). Table 2 shows an overview of our participants, their educational background, occupation, and experience with document accessibility. Participation was voluntary and not compensated.

ID	education	occupation	a11y exp.
E01	PhD chemistry	chemistry prof	BLV PhD student
E02	PhD chemistry	teacher BLV HS	see occupation
E03	MEd ph. & ch.	HS teacher	some, private life
E04	BSc chemistry	MSc chemistry	-
E05	HS	BSc biology	-
E06	BSc mech. eng.	BSc physics	SA
E07	HS	BSc physics	SA
E08	HS	BSc physics	SA
E09	HS	BSc physics	SA
E10	HS	BSc physics	SA
E11	HS	BSc physics	SA
E12	HS	BSc indust. eng.	SA
E13	BSc mechatronics	MSc mechatr.	SA
E14	BA music	MA music	-
E15	HS	law school	-

Table 2: Overview of think-aloud participants. HS = high school. All have HS chemistry knowledge, E01-E06 have had college level chemistry or above. E06-E13 are student assistants (SA) at the center for accessibility and are make literature from of maths, physics, and engineering accessible.

Procedure. Following established guidelines [1], we conducted a hybrid think-aloud while using *AutoChemplete*. Such hybrid sessions combine the benefits of concurrent and retrospective approaches, both capturing users thoughts as they occur, and allowing for semi-structured interviewing afterwards [1]. Sessions had an average duration of 55 minutes (SD=11) and were conducted via a video conferencing, were recorded and transcribed. Participants shared their screen to observe their interaction with *AutoChemplete*. Participants were introduced to the scenario that they are tasked with making a series of documents accessible, of which the structural formulas were already extracted and fed into *AutoChemplete*. Participants annotated one image after the other for about 20-30 minutes. Participants processed 12.2 images on average. Three participants completed only 5 images, whereas two managed to complete all 19 images prepared. After successful completion of the main task with *AutoChemplete*, participants were asked to briefly use a baseline version. In it, the model prediction (Figure 4a,b), and the autocomplete suggestions as seen in Figure 4c, were removed, leaving just the manual editor and the input image. As such our baseline version replicates the functionality of professional molecule-drawing applications like *ChemDraw*[17].

After the usage, the interviewer asked a series of semi-structured questions following a guideline for designing for motivation, engagement, and thriving in user experience ("METUX") [42]. It builds on the self-determination theory [14, 46], which explains human motivation and well-being based on fundamental psychological needs and has been frequently applied to HCI research [7].

4.3 Thematic Analysis

We conducted a thematic analysis following Braun and Clarke [9]. In a one-coder-one-interview scheme, we assigned a total of 2194 in-vivo codes to the interviews. The authors first aggregated these codes into a remaining total of 149 code groups. These groups were further iteratively refined into 26 higher order groups and ultimately seven themes that emerged during the analysis [18].

T1 - It Just Works. AutoChemplete enables participants to annotate chemical structural formulas, regardless of expertise. *"It is kind of easy"* (E03), for those with chemistry degrees, but even participants outside of STEM find it *"easy to use, there is a learning curve, [...] and while for the first few it took some time, it just works after that"* (E14). Consequently, users from both ends of the spectrum *"have a good feeling"* (E14) and are *"quite confident"* (E13, E03) in their results. Further, they find the *"speed"* at which they progress *"appropriate"* (E03) and think that *"suggestions contribute to this"* (E08). Even more importantly, participants that already perform such tasks as their student assistant jobs find the tool to be *"helpful"* (E12), and would like a comparable tool for their regular workflow around LaTeX math-mode. Despite, *"never having the situation to annotate a chemical formula before"*, E07 thought that the process *"worked well overall"*. In comparison to the baseline, which they "disliked strongly" (E15), users across the board voiced a strong "preference for the first" (E05), fully featured, version. In summary, for this first theme, we can report that participants have high satisfaction with the functionality and usability of *AutoChemplete*.

T2 - A Common Approach. We find users, independent of their chemical expertise, to portray the same behavior regarding their approach to annotation. As our editor closely resembles ChemDraw, a commercially available tool *"chemists often use"* (E03), one could assume that experts would tend towards using only this. The

[17] https://perkinelmerinformatics.com/products/research/chemdraw. For a web-based preview see: https://chemdrawdirect.perkinelmer.cloud/js/sample/

[18] We offer the guidelines and a codebook in the supplemental.

participant goes on to describe ChemDraw as *"a really powerful tool, maybe sometimes too powerful [...] to just add a figure to a paper"*. Our most expert participant E01 states that we have *"taken everything important out [of ChemDraw], everything important or essential is in this little builder, and everything else is just bells and whistles"*. However, users find suggestions beneficial, and *"like that you can have the side-by-side comparison because it makes it really easy"* (E01). *"If you just work with the molecules suggested here you basically only need to compare images."* (E03). Users notice the high success rate in finding *"the correct molecule very quickly"*, like a *"shortcut"* (E05) as previously outlined in 4.1. One of our most novice users, E15, thought aloud as follows: *"Well, there seems to be a sodium - was it? [confused about element symbols N and Na] - missing, so it can't be this one. Lets look at the next one. Okay, good, if its not this one then move on to the next and look at that."* Another says that *"their task was to simply check whether one image looks like the other"* (E06). As E11 puts it, *"only if the ones at the bottom are completely different, [they] would have to do it manually"*. On this basis we find users across the wide spectrum of expertise, from a chemistry professor (E01) to a law school student who did not take any chemistry courses in their last two years of high school (E15), to share a common approach of quick visual comparison.

T3 - Bond-by-Bond. However, we do find a stark contrast between how different groups perform this comparison. Novices regularly struggle with spatial orientation, due to rotation, mirrored molecules, or stereochemistry. As molecules are 3D structures, displayed in a 2D notation, there are multiple orientations, that may not look identical at first sight. *"Now it's mirrored again!"*, exclaimed E14, *"Why does virtually every one of those have to be mirrored?"*, summarizing a disdain of many less chemically-apt participants for spatial orientation tasks. E10 found this *"a bit confusing"*, and chose to *"rotate the molecule to compare in detail"*, as *"this makes it easier"*. In turn, experts feel at ease *"rotating things in [their] head to compare"* (E01). When the molecule in the editor was oriented in a different way than that in the input image, E03 recognized quickly that *"its okay the way it is"* and later explained that *"it's easy for me to imagine, because I dealt with stereochemistry in university. I had to practice this a lot!"* E13 voices concern about a more *"complicated molecule"*, where one needs to *"better check everything"*. During comparison of two molecules, E14 thinks out loud *"here needs to be a connection [bond], okay, now there should be two hydrogen atoms, and now [...]"*. Also E10 notices *"a bond left out"* or an *"atom missing"* from time to time. Their comparably low domain expertise is compensated well. The suggestions allow for convenient side-by-side view (c.f. T2). They spot slight differences - *"this one looks almost correct, this should be Cl [wrong atom in suggestion], I'll click on it to edit"* (E12) - and use this as a starting point for their corrections. *"One N, two H, [...]"* (E08) shows how novices approach molecular comparison in a bond-by-bond, and atom-by-atom fashion.

T4 - Names, Groups & Substructures. Expert users, however, compare differently. They *"appreciate that one gets all information for a [suggested] molecule"* (E05). This participant describes the expert mode of operation as: *"When I see a trivial name, I can often already say what molecule this is, what I may need to change. I didn't even need the structure, just the name"* (E05), hence chemists rarely need detailed comparisons. In other accounts, they could deduce molecular structures from *"the IUPAC convention, and just look at that to determine if there is a difference"* (E05). When it does come to visually comparing structures, even that is not done in the same way as the novices (T3). *"For the larger molecules, like lipids, I don't count each and every atom. One usually aggregates groups. I know the name of these groups, and this allows me for a lot of space in thinking"*, describes E05. Such liberty to approach the task is reflected by other expert participants, like E01, *"I didn't seem find myself wanting to do anything that I couldn't do"*. However, experts do appreciate the suggestions, *"as the quick way to go"*, however if the answer does not *"show up in the first few"*, experts *"just build it"* (E01). Experts simultaneously acknowledge the difficulties, novices face. *"If the system does not recognize it correctly, and well, imagine if you are not that confident, if you feel like a fish out of water, and really need to, well [laughs] count double bonds and carbon atoms, [...], it will be confusing."* (E02). One of our participants who is a teacher adds that this is the way how students are taught to read molecules even as early as in high school, *"search for the longest chain, or for functional groups, or ring structures, and then put it together piece by piece [...] however here, I just need to compare images"* (E03). In summary, our experts compare molecules via trivial or IUPAC names first, then move on to groups and substructures, and only rarely even see individual atoms or bonds.

T5 - Chaos in Chemistry. Both experts and novices alike criticize the lack of coherence in the way the input images look. Participant E03 remarked that *"it's interesting how bad some are, even with perfect eyesight I could have easily confused this [a very short bond] with a negative charge"*, and *"this part [double bond in one input image] looks really strange, almost as if there was a break - good that the system can deal with such issues"*. Beyond such errors in the input image, there further is a lack of standardization in chemistry, especially with regards to structural formulas. As E05 summarizes *"its always annotated a little bit differently"*. This encompasses *"simple differences like colors, but also the H's that are missing sometimes"*, says E15, and refers to the fact that often so-called skeletal formulas with implicit hydrogen notation are used. As described in T3, there is also the matter of orientation. Further differences include abbreviating simple groups like NH_2 or *"representing them with bonds"* (E10). Charges can be visualized in different ways as well. In high school, teachers *"make this obvious, as not to confuse students. We put a circle around charge indicators, so it's clear"* (E03), even though such conventions are not commonly accepted. Concluding, we find a lack of visualization standards, almost a *"chaos"* of forms, that novices need to *"get used to at first"* (E10).

T6 - Necessary Distance. *"From 20 students, 18 want their materials digitally"*, says participant E02, who teaches chemistry at a high school specialized on BLV students, *"however of those 18, everyone has different needs. Not all use a braille display, some only use a screen reader, others a combination."*. E02 further describes how their BLV school relies mostly on colloquial names, and IUPAC convention for digital output, and works with *"tactile magnet boards, with different shapes for atoms and bonds"*. However, during the pandemic *"lockdown made this all but impossible"*, as a sighted assistant would have to accompany the student in-person to build structures on this board. *"We tried to come up with alternatives, but*

students did not learn much during this time", said E02, and explained how digital tactile formats would be beneficial. While the more expert participants find *"SMILES intuitive to read"* and *"can make out functional groups in the string"*, they agree with the novices that it is *"cool to not need to know SMILES to annotate the images"* (E03). *"I haven't even looked at them once"*, adds E05, while E03 says they *"don't know how a blind chemist would work with the output"*. *"Since I annotated visually, I cannot imagine how a blind person does this,"* remarks E05. In the case of physics *"since we know it [LaTeX], there [...] is a good connection with the [BLV] students"* (E12). For the chemical formulas they find no such connection. *AutoChemplete* thus creates a perceptible distance between BLV students and the annotators. This distance, however, seems to be necessary, as otherwise novice and intermediate annotators would be unable to help.

T7 - In-flight Entertainment. Lastly, participants across the spectrum of expertise found *AutoChemplete* to be *"entertaining"* and *"nice to work with"* (E14). Frequently it was compared to *"a puzzle"* (E04) or a *"game"* (E01, E12) or described as *"fun"* (E05, E06, E12). Participants with rather high expertise emphasize how the tool *"struck the right balance where [...] it makes my brain work a little, but not so much that I am frustrated"* (E01), where *"it's not too difficult, [...] something you can do for a somewhat long time"* (E04). *"This labeling is nice to do, like on the side, it's low cognitive effort, as the suggestions are rather good. Maybe you can even listen to a podcast or music"*, suggests E03. Participant E01 even states: *"I could do this on a long flight, to pass time. A great way to keep myself distracted! [...] Could probably do a couple hundred before I get bored"*. However, also participants with less expertise describe it as *"putting stuff together, like a game, so satisfying and [...] casual, it shows you everything you need"* (E12). Participant E06 would *"specifically choose"* chemical literature to annotate in their student assistant job *"on purpose, because its so fun working with this tool"*, particularly compared to the baseline. Thus, we conclude that *AutoChemplete* has a certain entertaining character, making the annotation process that has previously been described as less than desirable at least somewhat more attractive.

4.4 Annotation Paths

In the think-aloud participants get an exact match in approximately 97% of cases. In Figure 6 we plot which paths users took to annotate. We find that the majority (57%) of molecules were annotated via accepting a suggestion. For the remaining 43%, the editor was involved in some capacity. We find 6% of molecules that were accepted exactly as the model prediction, without changes. In another 16% of cases, the initial model prediction is changed manually until it is finally accepted. For remaining 21% of molecules, we find users loading a suggested molecule into the editor. 12% of them are accepted immediately without further changes. These could have just been accepted as suggestions, however the user did load them into the editor (e.g. to rotate them). For the remaining 9 of these 21%, the user does make changes. Here we see a somewhat larger error rate in the acceptance from the editor (2% incorrect vs. 41% correct) as compared to those accepted from suggestions (1% vs. 56%).

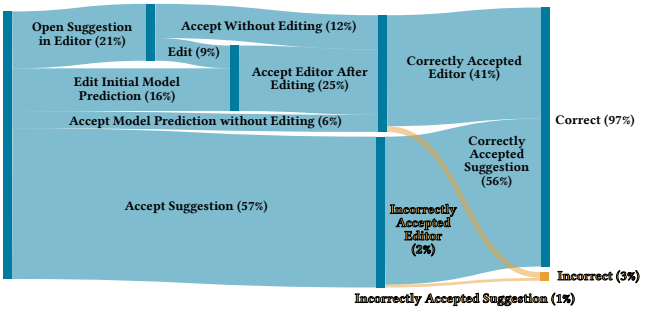


Figure 6: Annotation paths during the think-aloud sessions.

5 DISCUSSION

Overall, we find *AutoChemplete* to enable participants of every level of domain knowledge sufficiently well. The obtained levels of accuracy match current error rates at our center for accessibility and are generally remedied by its established four-eyes principle for annotation tasks. In the following, we discuss our results further.

5.1 Skill Balance of Autonomy and Competence

Motivational design research has long postulated trade-offs in skill support [28, 42, 45]. Foremost, existing research confirms the challenge to design for the needs of autonomy and competence, as posed by self-determination theory, in parallel. Hereby, an issue arises if the same system needs to appeal to both novices and experts simultaneously. Competence, a need for skill or mastery [14, 46], needs to be supported by guiding novice users. We elicit this as requirement R3 and report respective positive evaluation results of *AutoChemplete* in themes T1 and T3. However, such competence supporting design features could also thwart the need for autonomy of expert users. Autonomy thereby refers to a sense of volition or control [14, 46] and could be hindered by providing experts with help they do not need. *AutoChemplete* strikes a balance in such skill dependent situations, where both experts report that there was nothing they *"could not do"* (E01), but also novices are not frustrated and find that the tool *"helps to make it easy"* (E12).

5.2 Curb-cut Effects

Previous research and the interviews with C1-3 provided ample motivation to increase accessibility in STEM. Further, E04 mentioned that while there are trends to start requiring e.g. alt texts in some journals, there are still *"hundreds of papers published each month, maybe each week"* without such accessibility features. They further refer to *"all the old papers out there, going back 100, 150 years, millions in chemistry alone"*. However, sentiment in STEM is not always positive towards inclusion. Participant C3 described labs they applied to as a PhD student often voicing the attitude of *"you are welcome here, only if you can keep up with how we are working"*.

As making literature accessible is a high-effort task, another point of our evaluation was to identify curb-cut effects of *AutoChemplete* and the resulting annotated literature, to bring additional arguments. Experts pointed out that having *"the SMILES string of every structure in a paper, would be super useful and people would use that data"*, for instance to *"copy and paste"* molecules

from a paper into applications *"like ChemDraw without having to reproduce them"* (E01). Experts further note that by making literature accessible, one generally enables them to be *"machine searchable"* (E01), something that is already standard practice for *"company-internal databases"* (E02) e.g. of the *"pharmaceutical industry"* (E01). Such benefits apply to outlets that mandate accessibility annotations for publication, as well as ex-post annotation with tools like *AutoChemplete*. Further, participants in the educational sector note that a tool like *AutoChemplete* would help them to create *"better learning materials, assignment sheets or exams"* (E03). By using *AutoChemplete* in a standalone form, they *"can more quickly draw and export a structure, and also get suggestions for similar molecules to be used as examples"*, which they did *"maybe not think of before"* (E03). As theme T7 outlines, there seems to further be an entertaining character to *AutoChemplete*. This leads to participant E01 to suggest *"making a game out of it"*, where one can use images where the solution is known, to offer it as a learning tool to students, and *"mix in some unknown ones, so the students can help their [BLV] peers"*.

5.3 Generalizability

Beyond accessibility of chemical structural formulas, we see the possibility to generalize our approach to other human-AI interaction formats. Frequently, modern ML models [43, 58] approach high levels of accuracy. However, when exactness is required, they often cannot be used standalone. *AutoChemplete* combines such imperfect models with autocomplete in form of a solution space search, to enable humans to better perform their tasks. Such an approach implements the pattern of interactive labeling in ML and human-AI interaction guidelines [2, 3]. Potential candidates include the medical domain, where automating diagnosis and treatment selection seems far-fetched, however our proposed interaction pattern could allow for doctors to be suggested several treatment options to discover. Our approach favors problems, where there is a known, finite, but too-large to digest manually, solution space. We have visualized this general approach to such problems in Figure 7.

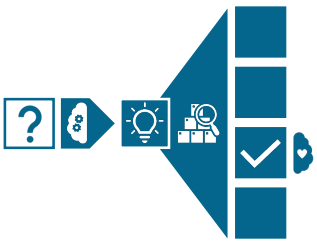


Figure 7: Proposed human-AI interaction pattern of *imperfect model + solution space search + human intervention.*

5.4 Limitations and Future Work

Our study is limited by its choice of application area, methodology, and participant demographic. While we outline how our approach may generalize, other areas may not deliver the preconditions for it to succeed. Accessibility annotations of plots or diagrams is a challenge, in which the solution space is not easily searchable. Following extant HCI research, however, we see transferability to e.g. mathematical formulas [17]. Future work could integrate such specialized tools into the document accessibility platform demanded in R5. Our qualitative findings demonstrate how we enable users of different backgrounds to work with *AutoChemplete*, however follow-up studies could quantitatively compare effects of support versus no-support in novices and expert groups, much alike [32].

More specifically, participants suggested two changes to the user interface. Firstly, they asked for the ability to "zoom, pan, and rotate the [...] image, just as [...] the editor" (E03). Hereby they referred to the fact that both source and autocomplete suggestion molecules were fixed in their orientation. With such an option to align the molecules, it "might be easier [...] to compare these side-by-side" (E13). Especially those without much chemical aptitude advocate that image rotation would help them with issues regarding spatial imagination. This suggestion not only supports R3, but is also trivial to implement. Thus we plan to follow it for the document accessibility platform that *AutoChemplete* will be part of.

Secondly, participants ask for an even larger emphasis to be placed on the suggestions: "swap the order of [user interface components], as I said, my first gaze went towards the suggestions immediately after the source image" (E08). Beyond changing the order of components, a revised version of *AutoChemplete* could propose a two-stage process. Our quantitative findings support the user's impressions that the correct result can be found in the top suggestions on most occasions (c.f. Figure 5). In a potential two-stage process, the user could initially be presented with only the source image alongside one or two suggestions. Input options for the user consequently are either accepting one of the suggestions, or opening either of them as a starting point in the editor and then being presented with the current interface of *AutoChemplete*.

6 CONCLUSION

Based on interviews we derived requirements for making chemical structural formulas accessible. In particular, we find that BLV chemists need exact, but individualized representations. However, accessibility professionals have little resources available to support. Thus they need to work quickly and with as little domain expertise required as possible. Lastly, such a tool needs to integrate into existing approaches. On this basis, we present *AutoChemplete*, an interactive labeling tool for chemical structural formulas, combining a state-of-the-art, yet imperfect, model, a similarity search in the solution space and human intervention in form of graphical annotation. We evaluate with a series of participants of varying levels of chemistry expertise. Participants think-aloud sessions and annotation paths reveal an overall similar approach of graphical comparison, however different styles of molecular analysis. With these insights we strive to inspire both future research in digital inclusion, as well as practical application of tools like *AutoChemplete* towards a world where STEM education is accessible for everyone.

ACKNOWLEDGMENTS

Funded by the Helmholtz Program Engineering Digital Futures, Topic 4 Knowledge to Action, Real-world Lab Digital Accessibility and Assistive Technology. We thank all anonymous interview & think-aloud participants for their support.

REFERENCES

[1] Obead Alhadreti and Pam Mayhew. 2018. Rethinking Thinking Aloud: A Comparison of Three Think-Aloud Protocols. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173618

[2] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35, 4 (Dec. 2014), 105–120. https://doi.org/10.1609/aimag.v35i4.2513 Number: 4.

[3] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300233

[4] Katrin Angerbauer, Nils Rodrigues, Rene Cutura, Seyda Öney, Nelusa Pathmanathan, Cristina Morariu, Daniel Weiskopf, and Michael Sedlmair. 2022. Accessibility for Color Vision Deficiencies: Challenges and Findings of a Large Scale Study on Paper Figures. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–23. https://doi.org/10.1145/3491102.3502133

[5] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2016. *Neural Machine Translation by Jointly Learning to Align and Translate*. Technical Report arXiv:1409.0473. arXiv. https://doi.org/10.48550/arXiv.1409.0473 arXiv:1409.0473 [cs, stat] type: article.

[6] Dávid Bajusz, Anita Rácz, and Károly Héberger. 2015. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics* 7, 1 (May 2015), 20. https://doi.org/10.1186/s13321-015-0069-3

[7] Nick Ballou, Sebastian Deterding, April Tyack, Elisa D Mekler, Rafael A Calvo, Dorian Peters, Gabriela Villalobos-Zúñiga, and Selen Turkay. 2022. Self-Determination Theory in HCI: Shaping a Research Agenda. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3491101.3503702

[8] Edward Bell and Arielle Silverman. 2019. Access to Math and Science Content for Youth Who Are Blind or Visually Impaired. *Journal of Blindness Innovation and Research* 9, 1 (2019), 1. https://nfb.org/images/nfb/publications/jbir/jbir19/jbir090101.html

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[10] Emeline Brulé and Gilles Bailly. 2021. "Beyond 3D printers": Understanding Long-Term Digital Fabrication Practices for the Education of Visually Impaired or Blind Youth. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3411764.3445403

[11] Joseph Chee Chang, Saleema Amershi, and Ece Kamar. 2017. Revolt: Collaborative Crowdsourcing for Labeling Machine Learning Datasets. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 2334–2346. https://doi.org/10.1145/3025453.3026044

[12] Djork-Arné Clevert, Tuan Le, Robin Winter, and Floriane Montanari. 2021. Img2Mol – accurate SMILES recognition from molecular graphical depictions. *Chemical Science* 12, 42 (2021), 14174–14181. https://doi.org/10.1039/D1SC01839F

[13] Maitraye Das, Thomas Barlow McHugh, Anne Marie Piper, and Darren Gergle. 2022. Co11ab: Augmenting Accessibility in Synchronous Collaborative Writing for People with Vision Impairments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3491102.3501918

[14] Edward L. Deci and Richard M. Ryan. 1985. *Intrinsic motivation and self-determination in human behavior*. Springer Science+Business Media, New York.

[15] Michael Desmond, Michael Muller, Zahra Ashktorab, Casey Dugan, Evelyn Duesterwald, Kristina Brimijoin, Catherine Finegan-Dollak, Michelle Brachman, Aabhas Sharma, Narendra Nath Joshi, and Qian Pan. 2021. Increasing the Speed and Accuracy of Data Labeling Through an AI Assisted Interface. In *26th International Conference on Intelligent User Interfaces*. ACM, College Station TX USA, 392–401. https://doi.org/10.1145/3397481.3450698

[16] Heather Devine, Andres Gonzalez, and Matthew Hardy. 2011. Making accessible PDF documents. In *Proceedings of the 11th ACM symposium on Document engineering (DocEng '11)*. Association for Computing Machinery, New York, NY, USA, 275–276. https://doi.org/10.1145/2034691.2034748

[17] Yancarlos Diaz, Gavin Nishizawa, Behrooz Mansouri, Kenny Davila, and Richard Zanibbi. 2021. The MathDeck Formula Editor: Interactive Formula Entry Combining LaTeX , Structure Editing, and Search. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–5. https://doi.org/10.1145/3411763.3451564

[18] Alfred T. D'Agostino. 2022. Accessible Teaching and Learning in the Undergraduate Chemistry Course and Laboratory for Blind and Low-Vision Students. *Journal of Chemical Education* 99, 1 (Jan. 2022), 140–147. https://doi.org/10.1021/acs.jchemed.1c00285

[19] Henri A. Favre, Warren H. Powell, and International Union of Pure and Applied Chemistry (Eds.). 2014. *Nomenclature of organic chemistry: IUPAC recommendations and preferred names 2013*. Royal Soc. of Chemistry [u.a.], Cambridge.

[20] Ann P. Gulley, Luke A. Smith, Jordan A. Price, Logan C. Prickett, and Matthew F. Ragland. 2017. Process-Driven Math: An Auditory Method of Mathematics Instruction and Assessment for Students who Are Blind or Have Low Vision. *Journal of Visual Impairment & Blindness* 111, 5 (Sept. 2017), 465–471. https://doi.org/10.1177/0145482X1711100507 Publisher: SAGE Publications Inc.

[21] Megan Hofmann, Kelly Mack, Jessica Birchfield, Jerry Cao, Autumn G Hughes, Shriya Kurpad, Kathryn J Lum, Emily Warnock, Anat Caspi, Scott E Hudson, and Jennifer Mankoff. 2022. Maptimizer: Using Optimization to Tailor Tactile Maps to Users Needs. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3491102.3517436

[22] Leona Holloway, Kim Marriott, and Matthew Butler. 2018. Accessible Maps for the Blind: Comparing 3D Printed Models with Tactile Graphics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173772

[23] Leona M Holloway, Cagatay Goncu, Alon Ilsar, Matthew Butler, and Kim Marriott. 2022. Infosonics: Accessible Infographics for People who are Blind using Sonification and Voice. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3491102.3517465

[24] Dorine in't Veld and Volker Sorge. 2018. The Dutch Best Practice for Teaching Chemistry Diagrams to the Visually Impaired. In *Computers Helping People with Special Needs*, Klaus Miesenberger and Georgios Kouroupetroglou (Eds.). Vol. 10896. Springer International Publishing, Cham, 644–647. https://doi.org/10.1007/978-3-319-94277-3_99 Series Title: Lecture Notes in Computer Science.

[25] Chen Jiang, Xi Jin, Ying Dong, and Ming Chen. 2016. Kekule.js: An Open Source JavaScript Chemoinformatics Toolkit. *Journal of Chemical Information and Modeling* 56, 6 (June 2016), 1132–1138. https://doi.org/10.1021/acs.jcim.6b00167

[26] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, and Evan E Bolton. 2021. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research* 49, D1 (Jan. 2021), D1388–D1395. https://doi.org/10.1093/nar/gkaa971

[27] Merlin Knaeble, Zihan Chen, Thorsten Schwarz, Gabriel Sailer, Kailun Yang, Rainer Stiefelhagen, and Alexander Maedche. 2022. Accessible Chemical Structural Formulas Through Interactive Document Labeling. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Georgios Kouroupetroglou, Katerina Mavrou, Roberto Manduchi, Mario Covarrubias Rodriguez, and Petr Penáz (Eds.). Vol. 13341. Springer International Publishing, Cham, 38–46. https://doi.org/10.1007/978-3-031-08648-9_6 Series Title: Lecture Notes in Computer Science.

[28] Tanner Kohler. 2022. Autonomy, Relatedness, and Competence in UX Design. https://www.nngroup.com/articles/autonomy-relatedness-competence/

[29] Nicholas Kong, Marti A. Hearst, and Maneesh Agrawala. 2014. Extracting references between text and charts via crowdsourcing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 31–40. https://doi.org/10.1145/2556288.2557241

[30] Jordan C. Koone, Chad M. Dashnaw, Emily A. Alonzo, Miguel A. Iglesias, Kelly-Shaye Patero, Juan J. Lopez, Ao Yun Zhang, Bernd Zechmann, Noah E. Cook, Mona S. Minkara, Cary A. Supalo, Hoby B. Wedler, Matthew J. Guberman-Pfeffer, and Bryan F. Shaw. 2022. Data for all: Tactile graphics that light up with picture-perfect resolution. *Science Advances* 8, 33 (Aug. 2022), eabq2640. https://doi.org/10.1126/sciadv.abq2640 Publisher: American Association for the Advancement of Science.

[31] Cheuk Yin Phipson Lee, Zhuohao Zhang, Jaylin Herskovitz, JooYoung Seo, and Anhong Guo. 2022. CollabAlly: Accessible Collaboration Awareness in Document Editing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3491102.3517635

[32] Jo-Yu Lo, Da-Yuan Huang, Tzu-Sheng Kuo, Chen-Kuo Sun, Jun Gong, Teddy Seyed, Xing-Dong Yang, and Bing-Yu Chen. 2019. AutoFritz: Autocomplete for Prototyping Virtual Breadboard Circuits. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300633

[33] Kelly Mack, Emma McDonnell, Dhruv Jain, Lucy Lu Wang, Jon E. Froehlich, and Leah Findlater. 2021. What Do We Mean by "Accessibility Research"? A Literature Survey of Accessibility Papers in CHI and ASSETS from 1994 to 2019. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3411764.3445412

[34] Joe R. McDaniel and Jason R. Balmuth. 1992. Kekule: OCR-optical chemical (structure) recognition. *Journal of Chemical Information and Computer Sciences* 32, 4 (July 1992), 373–378. https://doi.org/10.1021/ci00008a018

[35] M.B. McGrath and J.R. Brown. 2005. Visual learning for science and engineering. *IEEE Computer Graphics and Applications* 25, 5 (Sept. 2005), 56–63. https://doi.org/10.1109/MCG.2005.117 Conference Name: IEEE Computer Graphics and Applications.

[36] David Mendez, Anna Gaulton, A Patrícia Bento, Jon Chambers, Marleen De Veij, Eloy Félix, María Paula Magariños, Juan F Mosquera, Prudence Mutowo, Michał Nowotka, María Gordillo-Marañón, Fiona Hunter, Laura Junco, Grace Mugumbate, Milagros Rodriguez-Lopez, Francis Atkinson, Nicolas Bosc, Chris J Radoux, Aldo Segura-Cabrera, Anne Hersey, and Andrew R Leach. 2019. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Research* 47, D1 (Jan. 2019), D930–D940. https://doi.org/10.1093/nar/gky1075

[37] Mario Nadj, Merlin Knaeble, Maximilian Xiling Li, and Alexander Maedche. 2020. Power to the Oracle? Design Principles for Interactive Labeling Systems in Machine Learning. *KI - Künstliche Intelligenz* 34, 2 (June 2020), 131–142. https://doi.org/10.1007/s13218-020-00634-1

[38] Rahul Kumar Namdev and Pattie Maes. 2015. An interactive and intuitive stem accessibility system for the blind and visually impaired. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '15)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/2769493.2769502

[39] Abdulmumin A. Nuhu. 2014. Bioactive Micronutrients in Coffee: Recent Analytical Approaches for Characterization and Quantification. *ISRN Nutrition* 2014 (Jan. 2014), 1–13. https://doi.org/10.1155/2014/384230

[40] Srishti Palani, Zijian Ding, Austin Nguyen, Andrew Chuang, Stephen MacNeil, and Steven P. Dow. 2021. CoNotate: Suggesting Queries Based on Notes Promotes Knowledge Discovery. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. https://doi.org/10.1145/3411764.3445618

[41] Jungkap Park, Gus R Rosania, Kerby A Shedden, Mandee Nguyen, Naesung Lyu, and Kazuhiro Saitou. 2009. Automated extraction of chemical structure information from digital raster images. *Chemistry Central Journal* 3, 1 (Dec. 2009), 4. https://doi.org/10.1186/1752-153X-3-4

[42] Dorian Peters, Rafael A. Calvo, and Richard M. Ryan. 2018. Designing for Motivation, Engagement and Wellbeing in Digital Experience. *Frontiers in Psychology* 9 (2018), 797. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00797

[43] Yujie Qian, Zhengkai Tu, Jiang Guo, Connor W. Coley, and Regina Barzilay. 2022. *Robust Molecular Image Recognition: A Graph Generation Approach*. Technical Report arXiv:2205.14311. arXiv. https://doi.org/10.48550/arXiv.2205.14311 arXiv:2205.14311 [cs] type: article.

[44] André Rodrigues, Hugo Nicolau, André Santos, Diogo Branco, Jay Rainey, David Verweij, Jan David Smeddinck, Kyle Montague, and Tiago Guerreiro. 2022. Investigating the Tradeoffs of Everyday Text-Entry Collection Methods. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–15. https://doi.org/10.1145/3491102.3501908

[45] Richard M. Ryan (Ed.). 2019. *The Oxford Handbook of Human Motivation* (2 ed.). Oxford University Press, New York. https://doi.org/10.1093/oxfordhb/9780190666453.001.0001

[46] Richard M. Ryan and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55 (2000), 68–78. https://doi.org/10.1037/0003-066X.55.1.68 Place: US Publisher: American Psychological Association.

[47] Noureddin M. Sadawi, Alan P. Sexton, and Volker Sorge. 2012. Chemical structure recognition: a rule-based approach. In *Document Recognition and Retrieval XIX*, Vol. 8297. SPIE, Burlingame, CA, 101–109. https://doi.org/10.1117/12.912185 tex.ids= sadawi2012ChemicalStructureRecognition.

[48] Volker Sorge, Akashdeep Bansal, Neha M Jadhav, Himanshu Garg, Ayushi Verma, and M Balakrishnan. 2020. Towards generating web-accessible STEM documents from PDF. In *Proceedings of the 17th International Web for All Conference (W4A '20)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3371300.3383351

[49] Joshua Staker, Kyle Marshall, Robert Abel, and Carolyn M. McQuaw. 2019. Molecular Structure Extraction from Documents Using Deep Learning. *Journal of Chemical Information and Modeling* 59, 3 (March 2019), 1017–1029. https://doi.org/10.1021/acs.jcim.8b00669

[50] Cary A. Supalo, Mick D. Isaacson, and Michael V. Lombardi. 2014. Making Hands-On Science Learning Accessible for Students Who Are Blind or Have Low Vision. *Journal of Chemical Education* 91, 2 (Feb. 2014), 195–199. https://doi.org/10.1021/ed3000765 Publisher: American Chemical Society.

[51] Mingxing Tan and Quoc V. Le. 2020. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. Technical Report arXiv:1905.11946. arXiv. https://doi.org/10.48550/arXiv.1905.11946 arXiv:1905.11946 [cs, stat] type: article.

[52] Aniko T. Valko and A. Peter Johnson. 2009. CLiDE Pro: The Latest Generation of CLiDE, a Tool for Optical Chemical Structure Recognition. *Journal of Chemical Information and Modeling* 49, 4 (April 2009), 780–787. https://doi.org/10.1021/ci800449t

[53] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc., Long Beach, CA, 1. https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[54] Luis von Ahn, Shiry Ginosar, Mihir Kedia, Ruoran Liu, and Manuel Blum. 2006. Improving accessibility of the web with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. Association for Computing Machinery, New York, NY, USA, 79–82. https://doi.org/10.1145/1124772.1124785

[55] Yanan Wang, Ruobin Wang, Crescentia Jung, and Yea-Seul Kim. 2022. What makes web data tables accessible? Insights and a tool for rendering accessible tables for people with visual impairments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–20. https://doi.org/10.1145/3491102.3517469

[56] David Weininger. 1988. SMILES, a chemical language and information system. *Journal of Chemical Information and Computer Sciences* 28, 1 (Feb. 1988), 31–36. https://doi.org/10.1021/ci00057a005 Publisher: American Chemical Society.

[57] Robin Winter, Floriane Montanari, Frank Noé, and Djork-Arné Clevert. 2019. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. *Chemical Science* 10, 6 (2019), 1692–1701. https://doi.org/10.1039/C8SC04175J

[58] Sanghyun Yoo, Ohyun Kwon, and Hoshik Lee. 2022. Image-to-Graph Transformers for Chemical Structure Recognition. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Singapore, Singapore, 3393–3397. https://doi.org/10.1109/ICASSP43922.2022.9746088

[59] Beste F. Yuksel, Soo Jung Kim, Seung Jung Jin, Joshua Junhee Lee, Pooyan Fazli, Umang Mathur, Vaishali Bisht, Ilmi Yoon, Yue-Ting Siu, and Joshua A. Miele. 2020. Increasing Video Accessibility for Visually Impaired Users with Human-in-the-Loop Machine Learning. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3334480.3382821

[60] Lei Zhang, Yan Tong, and Qiang Ji. 2008. Active Image Labeling and Its Application to Facial Action Labeling. In *Computer Vision – ECCV 2008 (Lecture Notes in Computer Science)*, David Forsyth, Philip Torr, and Andrew Zisserman (Eds.). Springer, Berlin, Heidelberg, 706–719. https://doi.org/10.1007/978-3-540-88688-4_52

[61] Yu Zhang, Yun Wang, Haidong Zhang, Bin Zhu, Siming Chen, and Dongmei Zhang. 2022. OneLabeler: A Flexible System for Building Data Labeling Tools. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–22. https://doi.org/10.1145/3491102.3517612