

От теорий к моделям эмоций для искусственного интеллекта — основные методологические вопросы



А. В. Шиллер

Аспирант кафедры философии и методологии науки философского факультета Московского государственного университета имени М. В. Ломоносова
E-mail: shiller.a@gmail.com

Alexandra V. Shiller
PhD student, Department of Philosophy and methodology of science, Faculty of Philosophy of Moscow State University, Russia

За последние пятнадцать лет отмечается все возрастающий интерес к области исследований искусственного интеллекта (ИИ) и широкого круга проблем и задач, сопровождающих разработку программ «сильного искусственного интеллекта». Одной из таких задач является задача создания у ИИ аффективной или эмоциональной подсистемы, входящей в комплексную когнитивно-аффективную систему. Особенno важным требованием, выдвигаемым к таким когнитивно-аффективным системам, является их принципиальная формализуемость программными средствами, возможность встроить их в архитектуру искусственного агента. В связи с этим представляется важным проанализировать существующие модели эмоций человека, теоретические подходы к созданию эмоциональных систем и оценить возможности и ограничения для их формализации. Кроме того, разнообразие моделей и архитектур приводит к появлению ряда методологических трудностей и вопросов. Задача разработки интеграционной и формализуемой когнитивно-аффективной системы является исследовательским вызовом, предполагающим взаимодействие ученых из разных областей науки, а также учет практических запросов робототехники и отдельных разработчиков систем искусственного интеллекта. Отдельно стоит отметить роль философии и психологии как наук, предоставляющих теоретическую основу для построения моделей. В данном обзоре пред-

Как цитировать статью: Шиллер А. В. От теорий к моделям эмоций для искусственного интеллекта — основные методологические вопросы // Ценности и смыслы. 2018. № 4 (56). С.126–137.

ставлены некоторые теоретические подходы к созданию эмоциональных подсистем у ИИ, а также затронуты методологические вопросы, возникающие в связи с разработкой эмоциональной системы у ИИ.

Ключевые слова: искусственный интеллект, вычислительные модели эмоций, моделирование эмоций, методология ИИ, когнитивно-аффективная система, искусственные агенты.

Введение

В настоящее время большинство исследователей феномена эмоций стремится к созданию их интегративной модели. Одним из выдвигаемых к модели требований является ее способность встраиваться в архитектуру искусственного агента, то есть ее принципиальная формализуемость программными средствами. Эта задача является исследовательским вызовом, предполагающим взаимодействие ученых из разных областей науки, а также учет практических запросов робототехники и отдельных разработчиков систем искусственного интеллекта. В связи с этим особенно важно подчеркнуть роль философии и психологии как наук, предоставляющих теоретическую основу для построения моделей.

За последние пять лет отмечен быстрый рост количества вычислимых, доступных для оцифровки моделей эмоций и архитектур аффективных агентов. Такое разнообразие моделей и архитектур приводит к появлению ряда методологических трудностей и вопросов. В данной статье сделан обзор и проведен анализ некоторых теоретических подходов к созданию эмоциональных подсистем у искусственного интеллекта. В связи с этим целями данной статьи являются:

1. Обзор и анализ существующих теоретических подходов к моделированию таких процессов, как генерация эмоций и формирование эффектов эмоций.

2. Анализ причин отсутствия интегративной теории эмоций и, соответственно, вычислительной теории моделирования эмоций.

3. Определение набора вопросов, на которые должны отвечать теории, претендующие на универсальность, и роль ведущих теорий для моделирования генерации эмоций и формирования эффектов эмоций.

4. Проведение анализа данных, которые должны предоставлять теории.

С помощью моделирования эмоций исследователи пытаются воплотить аффективную систему в проектах ИИ, при этом многие исследователи рассматривают вопрос систематизации моделирования эмоций как

на уровне индивидуальных задач, так и на архитектурном уровне. Рейли [20] (изложил некоторые вычислительные задачи и сосредоточился на анализе подходов к расчету интенсивности и сочетания сходных эмоций. Брекенс с коллегами [4] исследовали использование абстрактных наборов теоретических формализмов для систематического сравнения некоторых существующих теорий когнитивной оценки. Канамеро обсуждал требования к дизайну аффективных агентов, сосредоточившись на роли эмоций в выборе действий [7]. Лиссети и Гмитрашевич определяли набор высокоуровневых компонентов эмоций, требуемых для вычислительных моделей, в рамках разработанной ими схемы «Представление об аффективном знании» [17]. Сломан с коллегами проделали огромную работу по исследованию структурных требований для разных классов приспособительного поведения, включая требования для разных типов эмоций внутри контекста их когнитивно-аффективной архитектуры [23]. Ортони с коллегами предложили высокоуровневую модель для архитектуры, которая эксплицитно моделирует эмоции, а также содержит информацию об аффективных состояниях и чертах как параметрах, оказывающих влияние на процесс [19]. Феллоус [11] рассмотрел требования к эмоциям и их необходимость для синтетических агентов с точки зрения нейронауки.

Такой широкий набор подходов к моделированию эмоций объясняется существующими методологическими проблемами в этой области исследований. Несмотря на большое количество уже созданных и существующих отдельно моделей эмоций и множества аффективных агентов и архитектур, отсутствует согласованность и ясность в отношении того, что означает понятие «модель эмоций» [15]. «Моделирование эмоций» может означать их динамическую генерацию в противовес моделям черного ящика, которые соотносят специфические стимулы и связанные с ними эмоции. Оно может означать генерацию выражений лица, поз и движений, означающих специфические эмоции, у искусственных агентов или роботов. Оно может означать моделирование эффектов от эмоций в процессе принятия решения и выбора поведения. А также оно может означать включение информации об испытываемых пользователем эмоциях в модель, созданную для обучающихся и предсказывающих решения систем.

Также наблюдается недостаток ясности относительно того, какие аффективные состояния моделируются. Термин «эмоция» в аффектив-

ных моделях может относиться к эмоциям самим по себе — коротким, преходящим состояниям; настроениям; сложным состояниям, например аттитюдам, и часто встречающимся состояниям, которые традиционно не относятся психологами к эмоциям (например, состояние сконфуженности, состояние потока).

Модели эмоций также сильно различаются в зависимости от того, какая из приписываемых им функций моделируется. Среди функций эмоций можно выделить управление целью и выбор цели, распределение ресурсов и координацию подсистем, координацию и коммуникацию между виртуальными агентами и людьми.

Одним из следствий такой терминологической неопределенности является тот факт, что когда мы начинаем читать статью про «моделирование эмоций», мы не знаем, чего ожидать, что будет описано в работе. Статья может описывать детали генерации выражений лица, синтез аффективной речи, модели черного ящика, сопоставляющие предметно-ориентированные стимулы и конкретные эмоции, вспомогательные для принятия решения формализмы, оценивающие альтернативные типы поведения. Более серьезное следствие — отсутствие основополагающих принципов при создании проектов по моделированию конкретных аффективных феноменов. При создании моделей эмоций должны быть получены ответы на ряд вопросов: Каковы вычислительные задачи, которые должны быть реализованы? Какие теории наиболее применимы для данной модели? Какие эмпирические данные необходимо найти в литературе?

Отсутствие согласованной ясной терминологии также затрудняет сравнение подходов в вопросах их теоретических оснований, требований для моделирования и их теоретических объяснительных способностей, а также их эффективности в конкретных приложениях.

Теоретические основания моделирования эмоций

В настоящее время при моделировании эмоций принято выделять две фундаментальные категории процессов: создание или генерация эмоций и эффекты или последствия эмоций. Для этих процессов подходят разные теории с различной степенью разработанности.

Рассмотрим процесс возникновения эмоций. Возникновение эмоций — эволюционирующий динамический процесс, который протекает между набором модальностей, взаимодействиями между ними и со сложной обратной связью. В то время как все модальности включены

в этот процесс, наше понимание этого феномена наиболее полно со стороны когнитивной модальности, и наиболее «живучие» модели возникновения эмоций содержат в себе *когнитивную оценку* (конечно, существуют и исключения [3; 5; 6; 25]). Также стоит отметить, что поскольку обладающие телесным воплощением агенты становятся все сложнее, будет возрастать необходимость включать в возникновение эмоций и некогнитивные модальности.

Наиболее влиятельные теории оценки для вычислительного моделирования — это те, которые созданы с использованием «вычислительно-дружелюбных» терминов. Первыми из них были теории, предложенные Ортони и коллегами, теперь упоминающиеся как ОКК-теория [13; 18]. Некоторые модели использую теорию оценки Фриджа [13; 14]. Теории оценки под авторством Шерера, Ройzman и Смита с Кирби также недавно начали использоваться в качестве основы вычислительных моделей [21; 22; 24]. Существует значительная область пересечения между этими теориями в плане используемых функций для оценки вызывающих эмоции пусковых триггеров (например, желательность, схожесть, ответственность агента), но также существуют и заметные отличия в предлагаемых классификациях эмоций и наборах независимых функций или переменных, используемых для характеристики стимулов-триггеров.

Рассмотрим процесс формирования эффектов эмоций. Для задач моделирования полезно разделять эффекты эмоций на две категории: первая категория — видимые, обычно очень драматические поведенческие и экспрессивные выражения эмоций; вторая категория — менее заметные, но не менее драматические выражения внутренних и когнитивных процессов. Теории эффектов эмоций не так хорошо разработаны, как теории о возникновении эмоций. В то время как основными подходами к возникновению эмоций являются аффективные модели и когнитивная оценка и существуют две теории: теория ОКК и теории когнитивной оценки внутри компонентного подхода. Для создания моделей эмоциональных выражений (эффектов эмоций) таких общих теорий просто не существует.

Одной из причин отсутствия интегративной теории является мультиmodalная природа эффектов эмоций. Комплексная теория эффектов эмоций потребует сбора множества выражений эмоций всех модальностей, включая ненаблюдаемые эффекты, связанные с физиологическим субстратом, который является посредником для возникновения наблю-

даемых проявлений, а также внутренние когнитивные эффекты. Такая теория должна, по сути, объяснить сам феномен эмоций. Поскольку исследования эмоций активно развивались в последние десятилетия, эта теория не может представлять собой «единую теорию эмоций». Кроме того, учитывая разнообразие процессов — посредников эмоций, создание такой единой теории кажется маловероятным. Акцентируя внимание на символических моделях и когнитивной модальности в существующих аффективных моделях, рассмотрим теории, подходящие для моделирования когнитивных эффектов эмоций.

В большинстве существующих аффективных моделей когнитивные эффекты эмоций не представлены в должной мере или разработаны слабо. Чаще всего они противопоставлены экспрессивным каналам выражения агента и его поведенческим решениям. На первый взгляд, моделирование «невидимых» когнитивных эффектов эмоций может казаться «убийственным» для аффективных моделей в построении агентной архитектуры. Может казаться, что прямое соотношение связи аффективных состояний с их экспрессивными проявлениями и поведенческими выборами является адекватным путем. Эксплицитные представления когнитивных процессов как препятствующих переменных между эмоциями и их экспрессивными и поведенческими выражениями позволяет так же, как и представление эмоций как медиаторов между стимулами и ответами, более гибко соотносить их внутри окружающей среды и поведения агента. Моделирование аффективных когнитивных связей также обеспечивает эффективные средства создания приемлемого поведения в более сложной окружающей среде и средства создания зависящих от личности аффективных акцентуаций, которые делают агентов более правдоподобными.

Было предложено несколько теорий для объяснения конкретных наблюдаемых когнитивных эффектов эмоций (КЭЭ), например: теория конгруэнтного настроению ответа [1], теория влияния эмоций на суждения и принятие решений [12], теория влияния эмоций на различные когнитивные процессы и процессы восприятия [10]. Теории акцентируют внимание на различных компонентах работы с информацией (например, внимание, память, автоматические и конструируемые процессы), и исследователи часто группируют аффективные влияния в разные категории, основанные на когнитивных структурах и процессах, которые находятся под влиянием.

В существующих вычислительных моделях исследованы две категории теорий: *теории распространяющейся активации* внутри семантической сети, представляющей память, и *параметрические теории*, которые утверждают, что аффективные факторы работают как параметры, индуцирующие переменные в когнитивных процессах (и, впоследствии, поведение). Теории распространяющейся активации ставят целью объяснить *аффективные основания (праймеры)*, например меньшее время ответа, необходимое для определения целей, которые аффективно-конгруэнтны вызывающему стимулу или обладают другой аффективной окрашенностью, и *конгруэнтный настроению ответ* — тенденцию предпочтительно вызывать в ответ схемы памяти, чья аффективная окрашенность совпадает с текущим настроением [2; 9].

От теорий к моделям

Прежде чем обсуждать и разрабатывать дизайн аффективных моделей более детально, полезно сформулировать, какие конкретно информацию и данные необходимо получить в идеальном случае для обеспечения развития модели. Такой прагматический подход позволит в дальнейшем оценить теории-кандидаты для определения наиболее подходящих из них задачам конкретного процесса моделирования.

В идеале теории должны предоставлять необходимые данные и детали для операционализации необходимых вычислительных задач, а также обеспечивать технические характеристики (спецификации) лежащих в их основе механизмов с помощью репрезентирующих их структур и процессов [4]. В связи с этим теории когнитивной оценки возникновения эмоций должны давать ответы на такие вопросы:

- Каково соотношение «стимул — эмоция» для интересующей области? Должно ли это соотношение быть реализовано непосредственно (область «стимул — эмоция») или с помощью взаимодействующих, не зависимых от области интереса функций (например, новизны, валентности, желательности, приятности, локуса контроля агента)?
- Как влияния внешних стимулов связаны с внутренними стимулами у эмоций-триггеров (вызывают в памяти или предвосхищают события и ситуации)?
- Каковы различные стадии процесса оценки, и какие функции входят в каждую стадию? От чего зависят эти стадии — от самой эмоции?

ции, индивида или контекста?

- Каковы различия и взаимодействия между различными процессами, обеспечивающими оценку?
- Какие факторы и как влияют на интенсивность эмоции?
- Какова природа «аффективной динамики» — скорости появления и распада эмоций? Как она изменяется в зависимости от самой эмоции, индивида и контекста?
- Могут ли возникнуть сложные эмоции с помощью процесса оценки? Если нет, как должны быть встроены в одну эмоцию потенциально конфликтующие между собой триггеры? Если да, как должны быть встроены эти сложные эмоции в когерентное, ясное и последовательное аффективное состояние?
- Какие когнитивные структуры необходимы для поддержки оценки (например, цели, ожидания, планы), и каков уровень сложности и природа этих структур?
- Какой уровень сложности необходим для конкретной задачи моделирования (например, сенсомоторный, схематический или концептуальный)?
- Какой должна быть структура понятия эмоции как такового в случае, когда оно является объектом, созданным в ходе процесса оценки? Какая информация должна быть представлена в этой структуре (например, интенсивность, триггеры, локус контроля, направление, цели)?

Теории когнитивных эффектов эмоций должны отвечать на такие вопросы:

- Какие когнитивные процессы и структуры находятся под влиянием конкретных эмоций, настроений, аффективных состояний или черт? Какова природа этого влияния? Каковы эффекты для динамических ментальных конструктов — ситуаций, целей, ожиданий и планов?
- Каковы содержание и организация подверженных аффекту структур долговременной памяти?
- Как когнитивная оценка затрагивается эмоциями?
- Каковы отношения между интенсивностью эмоции или настроения и типом и силой оказываемого влияния? Могут ли различные интенсивности эмоций или настроений вызывать качественно различные эффекты в когнитивных процессах?

- Являются ли отдельные эмоции переменными — посредниками эффектов эмоций (например, страх влияет на смещение внимания в сторону угрозы) или такими посредниками являются индивидуальные оценочные пространства [16]?
- Как и когда влияния сложных эмоций, настроений и черт соединяются?
- Существуют ли различные типы процессов, являющихся медиаторами влияний? Каковы взаимодействия и зависимости между этими процессами?
- Можем ли мы получить необходимые данные об этих внутренних процессах и структурах для создания вычислительных моделей?

К сожалению, существующие теории не позволяют получить ответы на все эти вопросы, хотя теории, направленные на объяснение механизмов влияний эмоций на сознание, не так тщательно разработаны, как теории когнитивной оценки. Теория когнитивной оценки не может дать ответы на вопросы с достаточным уровнем детализации для построения вычислительной модели. Фактически часто сам факт создания модели мотивирует к усовершенствованию связанных с ним психологических теорий. По одним вопросам существует показательный консенсус, например, касательно того, какие типы триггеров вызывают какие эмоции, а другие требуют значительных «догадок», например, вопрос о том, как противоположные эмоции должны соединяться и взаимодействовать. Более-менее хорошо определен аспект возникновения эмоций и эффектов эмоций относительно аффективной динамики: есть вычисления интенсивности эмоций и величины эффектов эмоций, их изменения во времени, а также соединение сложных эмоций и настроений, сложных влияний. Обычно в психологической литературе доступны только качественные описания этих взаимоотношений, хотя последние исследования выражения и распознавания эмоций и лицевых выражений позволяют получить многообещающие количественные данные [8].

Приведенные выше методологические вопросы выделены по трем причинам. Во-первых, они создают основу для оценки теорий-кандидатов. Во-вторых, они предоставляют основу для определения вычислительных задач, необходимых для воплощения в жизнь аффективных моделей. В-третьих, они помогают определить специфические аспекты протекания аффективных процессов, для которых должны быть разработаны более конкретные теории. Следующим шагом в моделировании эмоций

является обсуждение конкретных вычислительных задач, необходимых для воплощения аффективных моделей, но эта проблема является темой для последующих статей.

Заключение

В данной статье были рассмотрены основные теории, используемые для моделирования эмоций, конкретно — двух фундаментальных процессов: генерации эмоций и эффектов эмоций. Кроме того, были рассмотрены основные методологические вопросы, на которые должны отвечать подходящие для задач моделирования теории.

Можно сделать вывод, что в настоящее время для процесса генерации эмоций существует теория, претендующая на роль универсальной, в то время как для процесса формирования эффектов эмоций такой теории не разработано. Основными причинами отсутствия интегративной теории эмоций являются мультимодальность природы эмоций и ряд методологических трудностей: отсутствие теоретической и терминологической согласованности, отсутствие основополагающих принципов для создания моделей.

В заключение важно отметить, что, по мнению автора, одной из причин выбора исследователями и разработчиками именно этих двух процессов — генерации эмоций и формирования эффектов эмоций для моделирования — является наглядность и заметность процессов формирования и выражения эмоций. Однако это не означает, что с помощью моделирования только этих двух процессов получится раскрыть и передать всю сложность аффективной системы. Моделирование этих процессов является началом долгого пути моделирования сложной системы, похожей на человеческую, для тех случаев, когда исследователи ставят задачу раскрыть механизмы ее функционирования, а не только создать внешнюю видимость и схожесть. Для достижения правдоподобности и реалистичности агентов следует отойти от точных моделей и привнести в моделирование элементы ошибок и смешений, свойственных человеческой природе. Также автору кажется важным обратиться к философии — феноменологии и философии сознания, что позволит усилить теоретические основания моделей эмоций. Кроме того, достижение истинного понимания глубинных механизмов, лежащих в основе аффективной системы, и, как следствие, появление возможности перенести его в искусственных агентов, кажется более реальной задачей

с поддержкой философских теорий, хотя и более сложно реализуемой.

Литература / References

1. *Bower G. H.* Mood and Memory // American Psychologist. 1981. No. 36. P. 129–148.
2. *Bower G. H.* How Might Emotions Affect Memory / In S. A. Christianson (Ed.). Handbook of Emotion and Memory. Hillsdale, NJ: Lawrence Erlbaum, 1992. P. 121–145.
3. *Breazeal C., Brooks R.* Robot Emotion: A Functional Perspective. MIT Media Lab, 2005. P. 51–76.
4. *Broekens J., DeGroot, D., Kosters W.A.* Formal Models of Appraisal: Theory, Specification, and Computational Model // Cognitive Systems Research. 2008. No. 9(3). P. 173–197.
5. *Cañamero L.* Modeling Motivations and Emotions as a Basis for Intelligent Behavior. In W. Lewis Johnson, ed., Proceedings of the First International Symposium on Autonomous Agents (Agents'97), New York, NY: The ACM Press. 1997. P. 148–155.
6. *Cañamero L., Avila-Gracia O.* A Bottom-Up Investigation of Emotional Modulation in Competitive Scenarios. Second International Conference, ACII 2007 Lisbon, Portugal, September 12–14, ACII 2007, NY: Springer, 2007. P. 398–409
7. *Cañamero, L.D.* 2001. Building Emotional Artifacts in Social Worlds: Challenges and Perspectives. In Emotional and Intelligent II: The Tangled Knot of Social Cognition. Papers from the 2001 AAAI Fall Symposium. Technical Report FS-01-02, Menlo Park, CA: AAAI Press, 2001. P. 22–30.
8. *Cohn J. F., Ambadar Z., Ekman P.* Observer-Based Measurement of Facial Expression with the Facial Action Coding System / In J. A. Coan & J. B. Allen (Eds.) // The handbook of emotion elicitation and assessment. NY: Oxford, 2007. P. 203–221.
9. *Derryberry D.* Emotional influences on evaluative judgments: Roles of arousal, attention, and spreading activation // Motivation and Emotion. 1998. No. 12(1). P. 23–55.
10. *Derryberry D., Reed M. A.* Information Processing Approaches to Individual Differences in Emotional Reactivity / In R. J. Davidson, K. R. Scherer, H. H. Goldsmith (Eds.) // Handbook of Affective Sciences. NY: Oxford, 2003. P. 681–697.
11. *Fellow J. M. & Arbib M. A.* "Edison" & "Russel": Definitions Versus Inventions in the Analysis of Emotion. Who Needs Emotions? / Fellow J.-M., Arbib M. A. (Eds.). NY: Oxford University Press. 2005. P. 5–21.
12. *Forgas J.* Mood and judgment: The affect infusion model (AIM) // Psychological Bulletin. 1999. No. 117(1). P. 39–66.
13. *Frijda N. H.* The Emotions. Cambridge: Cambridge University Press, 1986. P. 52–78.
14. *Frijda N. H., Swagerman J.* Can computers feel? Theory and design of an emotional system // Cognition and Emotion. 1987. No. 1(3). P. 235–257.
15. *Hudlicka E.* What are we modeling when we model emotion? // AAAI Spring Symposium: Emotion, Personality, and Social Behavior. Vol. Technical Report SS-08-04, Stanford University, CA: Menlo Park; CA: AAAI Press, 2008. P. 52–59.
16. *Lerner J. S., Tiedens L. Z.* Portrait of the Angry Decision Maker: How Appraisal Tendencies Shape Anger's Influence on Cognition // Journal of Behavioral Decision Making. 2006. No. 19. P. 115–137.
17. *Lisetti C., Gmytrasiewicz P.* Can rational agents afford to be affectless? // Applied Artificial Intelligence. 2002. No. 16(7–8). P. 577–609.
18. *Ortony A., Clore G. L., Collins A.* The Cognitive Structure of Emotions. NY:

Cambridge, 1988. P. 55–87.

19. *Ortony A., Norman D., Revelle W.* Affect and Proto-Affect in Effective Functioning // Who Needs Emotions? / In J.-M. Fellous & M. A. Arbib (Eds.) NY: Oxford, 2005. P. 173–202.

20. *Reilly W. S. N.* Modeling What Happens Between Emotional Antecedents and Emotional Consequents: . Vienna: Proceedings ACE, 2006. P. 45–72.

21. *Roseman I. J.* A Model of Appraisal in the Emotion System / In K. R. Scherer, A. Schorr, T. Johnstone (Eds.) // Appraisal Processes in Emotion: Theory, Methods, Research. NY: Oxford, 2001. P. 68–91.

22. *Scherer K. R.* Appraisal Considered as a Process of Multilevel Sequential Checking // Appraisal Processes in Emotion: Theory, Methods, Research / In K. R. Scherer, A. Schorr & T. Johnstone (Eds.). NY: Oxford, 2001. P. 92–120.

23. *Sloman, A., Chrisley, R., & Scheutz, M.* (2005). The Architectural Basis of Affective States and Processes In J.-M. Fellous & M. A. Arbib (Eds.), Who Needs Emotions?. NY: Oxford University Press. 2005. P. 67–89.

24. *Smith C. A., Kirby L.* Consequences require antecedents: Toward a process model of emotion elicitation // Feeling and Thinking: The role of affect in social cognition / In J. P. Forgas (Ed.). NY: Cambridge, 2000. P. 21–36.

25. *Velásquez J. D.* An Emotion-Based Approach to Robotics: Paper presented at the IROS. Published in: Proceedings 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS199. 9. P. 235–240

FROM THEORIES TO AI`S MODELS OF EMOTIONS — KEY MYTHOLOGICAL ISSUES

Over the past fifteen years, there has been an increasing interest into the field of artificial intelligence (AI) research and a wide range of problems and tasks that accompany the development of "strong AI" programs. One of such tasks is the task of creating an AI's affective or emotional subsystem, that will be the part of a complex cognitive-affective system in the same time. It is especially important for a model to reach the required level of formalization competence. In this case it is important to analyze real possibilities of models and theoretical approaches, their capacities to the creation of emotional systems and assessment of opportunities and limitations for their formatting. In addition, the diversity of models and architectures leads to a number of methodological difficulties and issues. The task of developing an integrative and formalized cognitive-affective system is a research challenge involving scientists from different fields of science, as well as taking into account the practical needs of robotics and individual developers of artificial intelligence systems. Besides it is worth noting the role of philosophy and psychology as sciences, providing a theoretical basis for constructing models. This review presents some theoretical approaches to the creation of emotional subsystems in AI, and also touches upon the methodological issues arising in connection with the development of the emotional system in AI.

Keywords: artificial intelligence, computational models of emotions, emotion modeling, AI methodology, cognitive-affective system, artificial agents.